Vol. XXV FEBRUARY, 1941 No. 1

THE JOURNAL OF APPLIED PSYCHOLOGY

Edited by
JAMES P. PORTER, Ohio University
Athens, Ohio

TABLE OF CONTENTS

A. W. KORNHAUSER AND RICHARD S. SCHULTZ. Research on Selection of Salesmen 1

MARION A. BILLS. Selection of Casualty and Life Insurance Agents 6

ALBERT K. KURTZ. Recent Research in the Selection of Life Insurance Salesmen 11

O. A. OHMANN. A Report of Research on the Selection of Salesmen at the Tremco Manufacturing Company 18

JAY L. OTIS. Procedures for the Selection of Salesmen for a Detergent Company 30

THOMAS M. STOKES. Selection Research in a Sales Organization 41

GEORGE K. BENNETT AND SYDNEY ROSLOW. Extension of the Norms of the Columbia Vocabulary Test 48

T. W. FORBES. Psychological Applications to the New Field of Traffic Engineering 52

RALPH BEDELL. The Relationship between Self-Estimated and Measured Vocational Interests 59

MATTHEW LUCKIESH AND FRANK K. MOSS. The Effect of Line-Length on Readability 67

MILTON BLUM AND BEATRICE CANDEE. The Selection of Department Store Packers and Wrappers with the Aid of Certain Psychological Tests 76

WILLIS FISHER AND SAMUEL P. HAYES, JR. Maladjustment in College Predicted by Bernreuter Inventory Scores and Family Position 86

EDWIN E. GHISELLI. The Measure of Trade Name Familiarity 97

H. PHOEBE GORDON. Study Habit Inventory Scores and Scholarship 101

DON LEWIS. The Timbre Test in the Revised Seashore Measures 108

ALFRED C. WELCH. A Test of Competitive Advertising 113

ALBERT B. BLANKENSHIP. A Correction 124

News and Notes 128

Book Reviews 131

New Books and Pamphlets 138

RESEARCH ON SELECTION OF SALESMEN

ARTHUR W. KORNHAUSER
University of Chicago
AND
RICHARD S. SCHULTZ
Psychological Corporation

ON September 2, 1940, a symposium on Research in the Selection of Salesmen was held at Pennsylvania State College as part of the program of the American Association for Applied Psychology.* The objective of the meeting was to review the results of recent selection research in various fields of selling and to exchange ideas concerning procedures, accomplishments, and next steps. Five of the seven papers which were read at the meeting are printed in the following pages.

The reports assembled here reflect the new wave of interest in scientific study of salesman selection which has occurred in the past few years. The vigorous research efforts of the early 1920's had been allowed to wane. Now, however, a number of applied psychologists have taken up where earlier investigations left off—and with two marked advantages provided by the intervening years. One is the considerable development which has taken place in methods for measuring personality traits and interests by means of tests and question-

* The symposium was arranged by the Program Committee of the Section of Industrial and Business Psychology of the American Association for Applied Psychology. The Committee consisted of Millicent Pond, Chairman, Marion A. Bills, and Kinsley R. Smith. Arthur W. Kornhauser served as chairman of the meeting; the participants in the formal part of the discussion were Marion A. Bills, Albert K. Kurtz, O. A. Ohmann, Jay L. Otis, Vernon P. Scheidt, Richard S. Schultz, and Thomas M. Stokes.

blanks; the other is the psychologists' improved understanding of, and increased emphasis upon, clinical interviewing procedures in studying personality and motivation. On the side of the business community, at the same time, there has come a somewhat better appreciation of what psychologists can offer. The *need* for improved selection of sales personnel has been persistently felt in all but the most fortunate (or most blind) business organizations. In no other area of employment, in fact, has so lively an interest in improvement been apparent. With the interest, moreover, has come an increasing willingness to support research, but only in so far as the business man has been convinced that the research will yield results of *practical* value.

The studies reported here are important, however, not only as indications of the renewed research activity in the field of salesman selection; they likewise contain significant illustrations of the employment procedures that are being developed or refined, the scientific spirit characterizing the studies, and the positive accomplishments being achieved. Each of these matters deserves brief comment.

The selection methods emphasized in these studies fall into three groups:

1. *Standardized and evaluated application blanks (personal history ratings).* Weighted scores are derived to predict candidates' probable success or failure as indicated by such factors as age, marital status, education, membership in organizations, financial status, health, etc.
2. *Carefully planned "diagnostic" interviews and interviewers' rating scales.* The traditional interview is extended and organized to obtain more adequate appraisal of candidates in relation to job demands. Varying degrees of standardization of the interview are being tried; likewise different guiding hypotheses for the analysis of personality.

3. *Standardized tests and question-blanks.* These are usually directed at the following qualities:
   a. *Interests and attitudes.* Questions concerning likes and dislikes, preferences, etc., commonly found in the interest and attitude scales.
   b. *Personality traits.* Questions aimed at indicating emotional stability, extroversion, self-sufficiency, aggressiveness, sociability, etc.
   c. *Mental alertness.* Ability of a salesman to understand his job and to benefit from special instructions and training material may sometimes be shown by intelligence tests; they may also indicate promotability.
   d. *Special knowledge and skill.* For some sales organizations it is essential that applicants possess basic information (*e.g.*, of finance, chemistry, mechanics, etc.) which can be measured by tests.

It is especially noteworthy that practical studies do not limit themselves to psychological tests. While the field of personality testing has great promise, it is clear that psychologists will need to continue utilizing and perfecting selection procedures of the other types listed rather than placing sole reliance upon tests.

With respect to scientific procedure in selection research, "good practice" has become fairly well established among professional industrial psychologists. There is general agreement, for example, on the desirability of such elements of method as the following, which are well illustrated in the studies reported:

1. The tests were carefully selected or especially designed for the organization, after a study of the job and its modifying conditions affecting success or failure.
2. Precise studies were made of measures of success as a basis for evaluating the effect of selection instruments.
3. Each selection procedure was tried out, carefully ana-

lyzed, and evaluated by reference to the criteria of success.

4. The effective procedures were then combined into a practical selection program, for further check and improvement prior to final adoption.
5. Periodic and continuous studies were instituted to check, improve, and adapt the selection methods as conditions change.

The last two points in this list are especially to be underscored. The essential final checking process and the later re-checks are probably the most neglected phases of scientific selection research. It cannot too often be emphasized that procedures carefully "tailored" to show good results for an experimental group of employees or applicants are not thereby demonstrated to be useful with the subsequent run of applicants. There is bound to be a certain conflict here between scientific caution and immediate practical demands. The psychologist can never wait until *all* the evidence is in, until he has indisputable proof of his selection instruments. But just because this is the case, and because the pressure to *use* the new techniques is so insistent, the psychologist must be everlastingly on guard against claiming too much too soon.

Finally, a word about the accomplishments of research on salesman selection, as reflected in the reports which follow. In the first place, it is abundantly clear from these and other investigations that effective selection procedures must be worked out in relation to the particular type of selling; there are at present no *general* "sales ability" tests. Where psychologists have carried on thorough and continuous research, however, to develop tools which may aid management in selection of specific sales personnel, definite results of value have been achieved.

In the second place, the accomplishments at best are substantial but not spectacular. They constitute useful, practical improvements in methods; they considerably increase the probabilities of selecting successful salesmen. The need for

caution in "selling" what the research can accomplish is apparent, however, to anyone who examines findings critically. No short-cut and "sure fire" devices are in sight. What can be offered is a confident expectation that long-run continued research in a sales organization will gradually yield sounder selection results than those produced by traditional procedures.

In conclusion, it is to be kept in mind that improved methods of selection must be built into procedures and policies already in operation in a company and they must be combined with sound programs of training, supervision, and employee relations generally, if best results are to be obtained. Even where a psychologist embarks upon a quite circumscribed selection study, he almost inevitably encounters related problems toward the solution of which his research methods can contribute. Certain of these relations are especially clearly stated in the last of the papers which follow.

SELECTION OF CASUALTY AND LIFE INSURANCE AGENTS

MARION A. BILLS

Assistant Secretary, Ætna Life Insurance Company

THE work on which I am reporting started in January 1933, and has continued up to the present time. A report of the first study was given at a meeting of this organization in Columbus, in 1937, but since I cannot compliment myself by assuming that many of you either heard or remember that report, I will very briefly outline the general set-up of the work.

The subjects for this experimentation have been the men attending the casualty insurance schools at the Ætna. There are seven of these a year, each lasting five weeks. The men vary widely both in age and in experience in the business. They come at their own expense and we are especially anxious to have those who will return to our various agencies as commissioned salesmen. The group, therefore, as you can see, has been a pre-selected one in that they have had sufficient interest in selling to come to the school and they likewise have had the training of the school.

We have used as test material, Strong's Vocational Interest Blank, Bernreuter's Personality Inventory, a personal history blank, and in some instances, a mental alertness test. The most useful of the tests have been the Strong's Vocational Interest Blank and the personal history blank.

As our criterion of success, we considered using actual production, but found on account of the peculiarities of the casualty business and the large amount coming to us through brokers, that this was not feasible. Therefore, our criterion of success has been the rating of the men by the managers to whose agencies they have returned. They were rated in

terms which came natural to the manager: outstanding success; success; still in business, will probably succeed; still in business, will probably fail; out of the business.

For a study made in January of this year, we had test material on 1,393 individuals who had been in the field a full year following the school. Of this number we got satisfactory success criteria on 700, just about half, which is probably about a typical falling off of the test group. Reasons for unsatisfactory or unreliable criterion of success came from those who, for example, had become special agents, had returned to a bank or brokerage house where we know the amount of insurance we get from the broker but have no way of telling how much an individual is producing, and, especially with the younger group, the fact that they had accepted salaried positions with our company or with other companies.

For our study we have used the score on two of the vocations of Strong's Vocational Interest Blank, life insurance and real estate salesmen. We have called an A, +3; a B-plus, +2; B, zero; B-minus, −2; and C, −3. That is, the total score for the two occupational scales goes from +6 to −6. In a previous study[1] it was found that for the total group the percentage of failures ran from the high score to the low score as follows: 22, 28, 42, 53, 76. The 1940 study gives just about the same ratio. However, it became evident from the first that the proportion of success and failure varied definitely with the different ages and therefore in this larger study three age groups have been separated: 24 and under; 25 to 29; and 30 and over. The 700 were divided fairly evenly among the age groups.

Although the percentage of failure differed with the age groups, in each age group the percentage of failure decreased with the increased score, with the exception of those under 24 where the score of +5 or +4 was more advantageous than a +6. The reasons for this will be discussed later.

[1] M. A. Bills, "Relation of Scores in Strong's Interest Analysis Blanks to Success in Selling Casualty Insurance," *J. Appl. Psychol.*, 22, 1938, 97–104.

Our second step was the scoring of the personal history blank. At first we used the weights suggested by the Life Insurance Sales Research Bureau with good results, but came to feel from our investigation that different age groups should be scored on different items. For example, with those under 24 we have found that there is a difference in proportion of success according to education, a college graduate being most likely to succeed, a high school graduate next, and least likely one having one to three years of college. By the time one is 30 and over other factors have apparently crowded out the predictiveness of education. For those under 25, too few were married or had dependents to make a scoring possible, but for those over 30, a married man with one or two dependents was more apt to succeed than a single person with no dependents, and much more apt to succeed than the person who had been divorced or separated, regardless of the number of dependents. As a second example, to carry any life insurance was a good indication for a person under 25, although $5,000 or over was better than a lower amount, but for a person over 30, $5,000 and over is good, none ranks next, and $1,000 to $5,000 gives a very poor rating. I believe that this is understandable in that when one has reached the age of 30 if one can get insurance at all and cannot afford to carry more than $5,000, it indicates a poor financial standing.

Considering the combined ratings on Strong's Vocational Interest Blank and the personal history blank, we find for those under 25 that the chances of success are 3.3 times as great if one scores high as they are if he scores low; for ages from 25 to 29, 2.4 times as great; and for a person 30 or over, 2.3 times as great. On the surface this does not seem any better prediction than the prediction quoted previously on the Strong Blank alone. The increase in value is that with the addition of the personal history, the proportion of people who fall definitely in the upper and lower predictive ranks is increased, and the number on which the prediction is indefinite, that is approximately 50 per cent, is decreased.

We have noted that for those under 24, a score of +4 or +5 was better than a score of +6. In the first study it seemed to us it was probably simply a chance variation but since it has persisted, we have studied it more carefully. Our conclusions have been that for a certain number of these men there was unconscious bluffing and that when this occurred, the probability was that they overcompensated and therefore scored a +6 on sales items. We made a special study of 25 individuals, picked purposely from an older group to preserve our younger group for the try-out, who scored +6 and who were rated as highly successful and 25 who were rated as failures. We found that there was a difference in their scores on certain other occupations. For example, the successful salesmen tended to score lower as purchasing agent and clerk than did the unsuccessful. On the other hand, there was a definitely larger percentage of successes who scored B as lawyer than of those who were unsuccessful. This bore out a feeling that we have had for some time that a secondary interest in law was advantageous.

By introducing a scoring scheme based on these extraneous occupations in our under 25 age group, we were able to spread those receiving +6 from 31 per cent failure to 55 per cent failure, and by scoring the entire group we were able to spread them in ratios from high to low score failures of 25, 33, 40, 63, and 78 per cent.

This method of scoring is sufficiently complicated so that we probably will not present it to our field, but, by keeping it in mind, one can at times warn the manager employing the man against what would seem to be unconscious bluffing or momentary enthusiasm for that particular type of work.

There was some question whether these same data could equally well be used for our life insurance agents as for our casualty agents. We offered the service of scoring Strong's Vocational Interest Blank to the field about a year and a half ago with the very definite warning that our data were based on casualty agents and might not be applicable to their group.

At the present time we have dependable but limited results on consistency of service and production for 220 life agents. To quote one age group as an example, 70 men under 24 with no previous experience in selling to whom the tests were given and who were later contracted and who have been selling for over one year, for those with scores on the Strong's Interest Analysis from +6 to +1, we had a 41 per cent cancellation, while from those scoring from 0 to −6 we had 69 per cent cancellation. The average production of those remaining who scored high on this blank was $44,000, while for those who scored low it was only $34,000. The combination of the scores on Strong's Vocational Interest Blank and the personal history as developed for the casualty agents and applied to the life agents, gave an average production running from those making high scores to those making low, of $51,000, $43,000, $33,000, and $28,000. That is, the Strong's test alone gives us a fair prediction of whether or not the man will remain in the business and some differentiation in production, while combined with the personal history it gives a much more definite prediction of production if the person remains with the company. These results are probably as good as one would expect when one considers that the test was definitely used as a selection aid and that therefore the number of people contracted with low scores was small; also there had been some selection against the test, since we had urged the agents to take on low score people only if all other items seemed especially desirable. At the present time we are accumulating data with which we can check past results.

RECENT RESEARCH IN THE SELECTION OF LIFE INSURANCE SALESMEN*

ALBERT K. KURTZ

Life Insurance Sales Research Bureau, Hartford, Connecticut

ALL of the research on the selection of salesmen that has been done under the auspices of the Life Insurance Sales Research Bureau has been done on the selection of *life insurance* salesmen. Because of the nature of our organization, we have not been much concerned with the broader problem of the selection of salesmen in general, and we do not know whether or to what extent our tests may be applicable outside of the life insurance industry. Consequently, the specific tests which we are using will probably not be of particular interest to you, but you may be interested in the general nature of these tests, the methods used in their construction and validation, the procedure we used in determining how to combine the scores, and the results obtained from their use (both on the original group and on an alien population).

At the present time, most of the large life insurance companies are using a test booklet we developed which is called the "Aptitude Index for Life Insurance Salesmen." This is in two parts and is really two separate tests: (I) a Prediction Scale based on personal history items and (II) a test of Personality Characteristics.[1]

* Read at the Annual Meeting of the American Association for Applied Psychology, September 2, 1940, at Pennsylvania State College.

[1] The Prediction Scale based on personal history items (Part I of the Aptitude Index) was developed by Albert K. Kurtz while working under Rensis Likert, formerly head of the Research Department of the Life Insurance Sales Research Bureau. The test of Personality Characteristics (Part II of the Aptitude Index) was developed for the Life Insurance Sales Research Bureau by Arthur W. Kornhauser, Associate Professor of Business Psychology at the University of Chicago.

A year ago at the joint meeting of the Psychometric Society and the American Psychological Association, I described the methods used in developing the Prediction Scale. It is based solely upon personal history items of the type found on application blanks, such as occupation, number of dependents, living expenses, etc. At that time I presented figures showing the effectiveness of that Prediction Scale, both on the original group of 10,111 salesmen who were used in its development, and on two alien populations of 743 and 878 life insurance salesmen. Therefore, I shall not go into further detail today regarding the construction and validation of the Prediction Scale.

In 1932, Dr. Kornhauser began work on the measurement of personality characteristics of life insurance salesmen by constructing a number of tests designed to measure various traits which were believed to be important factors in determining the success of men in this field. A few examples of such traits or characteristics are: self-confidence, self-assertiveness, dominance, emotional adjustment, and understanding of people. In the course of this experimentation, he constructed eight subtests containing about 500 test items or questions which were intended to measure 38 different characteristics similar to the examples just given. These subtests (or revisions of them) were answered by a total of approximately 1,000 men who were already in the life insurance business and whose degree of success or failure was a matter of record. As a result, he found that four of these subtests discriminated with a fairly high degree of accuracy between the more successful and the less successful in these groups of men already engaged in the business of selling life insurance.

Because these questions were measures of personality, interests, attitudes, and beliefs—rather than measures of ability—the next step obviously was to try out these four types of questions on life insurance agents just being hired in order to find out whether the mental characteristics which caused high test scores were the cause or the effect of success in selling life

insurance. In other words, we knew that these four subtests would work on men already in the business, but we did not know whether they would also work when answered by men at the time of hiring. Consequently, these questions were printed in a booklet and they were answered by 1,433 new life insurance salesmen hired by the Sales Managers of 24 life insurance companies. From this group, 304 salesmen were selected for further study; 211 of these men were 26 years of age or older, and 93 were 25 or younger. These 304 salesmen formed a homogeneous group: each one of them answered the test questions at the time he was hired, each of them actually went into the business of life insurance selling on a full-time basis, and on each of them we had sufficient information to classify him as to the degree of success he had attained. I might mention, however, that no salesman was excluded from this group because of either a high or a low test score, nor because of either a high or a low sales record.

Because of the fact that the Prediction Scale (based on personal history items) had already shown that it had a high degree of validity, we were concerned not only with whether this new test would prove valid, but we also wanted to find out whether or not a combination of these two tests would prove of greater value than either one alone; and, if so, how the two should be weighted. There was reason to believe that the personal history Prediction Scale was less valid at the younger ages than at somewhat older ones; and consequently separate tabulations were made for a number of different age groups. Finally, the data were divided into two age groups: 25 and younger, and 26 and older. Subsequently, these two age groups were handled separately. One of these four subtests proved to be of no value when thus tried out on men at the time of hiring; the other three were valid and form the present test of Personality Characteristics.

Multiple correlation coefficients between the criterion and the two tests (Personality Characteristics and the Prediction Scale) were computed for both age groups. These multiple

correlation coefficients were sufficiently higher than the zero order correlation coefficients to justify us in using both tests. However, we did not wish to use partial regression weights if we could avoid them. Consequently, by experimentation with various weights, we discovered that almost as high a correlation could be obtained by the use of simple weights as could be obtained by the use of the technically more accurate partial regression weights.

It was found that at the very young ages (that is, 25 and younger) the Personality Characteristics test should be given about one and one-half times as much weight as the Prediction Scale (based on personal history items) in order to predict a young man's likelihood of success with the greatest accuracy. On the other hand, at the relatively older ages (26 and older), the situation is exactly reversed. Consequently, it can be seen that it is very easy to explain the method, and very simple to combine the scores on the two parts in order to obtain a combined rating. However, even this step has been simplified by the use of a pair of diagrams showing the combined rating for every one of the possible combinations of scores on the two tests.

When the two tests are scored and the scores are combined in this manner, the result is a final classification into one or another of five categories: A, B, C, D, or E. The validity of the combined rating is in the neighborhood of +.40, as far as predicting actual sales is concerned. However, this combined rating also has a definite validity in predicting survivorship, since men who rate A have approximately twice the chance of remaining in the life insurance business for a year that men with ratings of E have. When the two factors of survival rate and sales are combined, the result is as shown in Chart 1. This chart assumes that equal numbers of men are hired in each of the five categories (A, B, C, D, and E) and shows (in relation to the average) the total sales made during their first year by the life insurance salesmen who remain in the business one year or longer. Any sales made by men who drop out

during their first year are eliminated from this picture. However, if these sales were included, the figures would not be appreciably altered.

Chart 1 is based upon 211 of the 304 men I mentioned earlier. These 211 are the men who were 26 years of age or older at the time they were hired. The validity of the Aptitude Index for the 93 men who were aged 25 and younger is approximately the same—the differences between the extreme groups are just as great, but there is a little more irregularity in the intervening categories, due to the small number of cases.

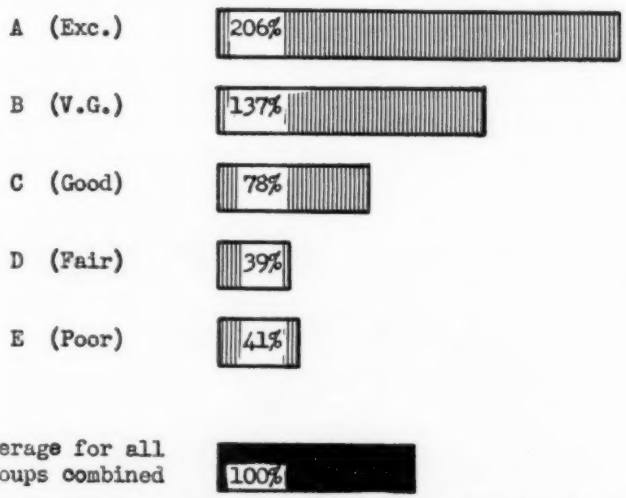

CHART 1. Aptitude Index: Survival Rate and Sales combined, age 26 and older (N = 211).

It is upon the basis of results such as those just shown in Chart 1, coupled with the desire of life insurance companies to improve the selection of their salesmen, that the combination of these two tests, known as the Aptitude Index, has been adopted by over 70 of the 106 life insurance companies that are members of the Life Insurance Sales Research Bureau.

Since its use in the actual hiring of new life insurance agents, there has been a marked decline in the number of men hired with combined ratings of D or E, but the predictive value of the Aptitude Index remains about the same as it was in this earlier group. This is shown in Chart 2. Chart 2 is based upon 372 life insurance salesmen who are the first of a larger group on whom we expect to get data during the next few months. Some of these salesmen have had previous life insurance selling experience, some are women, all ages are included, and there are various other factors that make this a relatively heterogeneous group. Later, we expect to analyze these records in terms of the various factors making for this heterogeneity, but for the present purpose, I felt it best to include everyone on whom we had a record up through last Wednesday (August 28, 1940). It will be noted that these two tests have proved to be valid, since the results on this alien population are just about the same as those on the earlier group of 211 men.

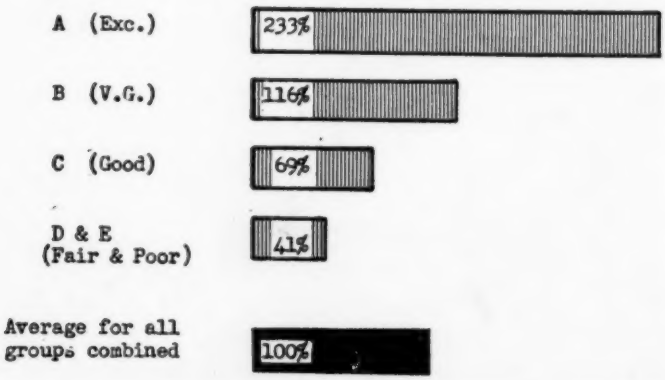

CHART 2. Aptitude Index: Survival Rate and Sales combined, all ages (N = 372).

I have no data to show that these specific test items, or even the same general type of test, will give results with any valid-

ity whatever in other occupations, even though they might superficially appear to be closely related to life insurance selling. However, I do feel that our research has demonstrated that it is possible to construct tests and to validate them in such a manner that life insurance selling ability can be predicted with a fairly high degree of accuracy. It seems to be quite likely that if similar procedures were applied in other occupations where valid criteria are available, and if a sufficiently large body of data were used in deriving and validating the measures, equally good results might be obtained in the prediction of selling ability in other fields.

A REPORT OF RESEARCH ON THE SELECTION OF SALESMEN AT THE TREMCO MANUFACTURING COMPANY

O. A. OHMANN
Western Reserve University

INDUSTRIAL psychology has made most headway in the large, well-established business organizations, where a long-time research point of view toward personnel problems has been developed, and where there are enough employees in any given occupation to make quantitative measurement both possible and profitable. In Cleveland there are 50 firms employing 500 or more workers, 108 firms with 200–500 workers, and 157 firms with 100–200 employees. This report may suggest some possibilities for psychological research in a relatively small organization, provided a continuing personnel research program has been established. A long-time approach to the problems of selection is imperative here because of the small samples available for study at any one time. Research attacks on various fronts may be made simultaneously, and measurements and records accumulated over a period of years for later validation. This report will be in the nature of an outline of the program of research on the selection of senior salesmen which has been initiated at Tremco Manufacturing Company. Progress made in various areas of study will be indicated and tentative findings summarized.

The Tremco Manufacturing Company operates plants in Cleveland and Toronto for the manufacture of materials used in the construction and maintenance of buildings—such as materials for caulking, pointing, waterproofing, roofing, painting, flooring. They employ 75 salesmen located in territories in the eastern half of the United States and all of Canada. These men call on plant superintendents, schools, apartment

owners, hospitals, large property owners, and on contractors and architects. They do not as a rule sell to individual home owners, nor to dealers or jobbers. Each sale is creative in the sense that it requires the diagnosis of a particular problem on a building, and the specification of materials and methods for making the repair. The importance of an effective sales force in such a business and the high cost of turnover indicated that the problem here was one of improving methods of selection and training. The research approaches which have been made to the problem of selection may be outlined as follows:

1. *Study of criteria.* The problem here was to determine the best measure of performance of salesmen, so that predictive items and instruments could later on be validated against this criterion. Fortunately territories had been established on the basis of equal sales potentials, as determined by various business indices. The experimental group for the study of criteria consisted of 30 salesmen who had been with the company for more than two years, and whose records were sufficiently complete for the year 1937 to permit the intercorrelation of the following measures of job performance:

Sales volume for 1937
Average number of calls per day
Number of years worked at Tremco
Salesman's net commission earnings for 1937
Average number of sales per month
Average size of order
Average number of new accounts per month
Average sales volume per year for length of time employed
Sales volume for the first six months on the job
Trend of sales volume over a period of years
Amount of allowances to customers
Amount of returned merchandise
Classes of trade called on
Classes of products sold.

This study resulted in establishing the "salesman's net commission earnings" as the best single criterion. The investi-

gation also yielded other interesting by-products. For example, "number of calls per day" correlated negatively with all other measures of success, including the "number of sales." Salesmen on the job a long while tend to open very few new accounts, although there is a possibility of doing this. Consistency of performance of salesmen from year to year was shown by a correlation of .92 between "1937 sales" and "average sales volume for length of time employed." However, "sales volume during the first six months" did not correlate highly (.44) with later annual sales volume. This last finding suggested the advisability of a more thorough investigation of the relation between sales during the first months of employment and later success.

In order to determine whether sales during the first few months might be used as a basis for early elimination of men who are likely to fail later on, the records of a group of 65 salesmen were studied. Cumulative sales volume for the first, second, third, and fourth months were correlated with year-end sales volume. Since these correlations range between .20 and .30 it was concluded that too much stress had been placed on the sales volume of new men. Subsidies to new men (in the form of overdrafts permitted) were entirely uncorrelated with year-end sales volume. Further analysis showed that sales volume during the first four months could be used to predict the most serious ultimate failures, but that average or high sales during the first few months did not indicate a corresponding degree of later success. Company policies with reference to subsidizing new salesmen have been adjusted in line with these findings.

2. *Study of the application blank and personal history data.* A four-page application blank with 31 items had been in use for eight years. Each of these items was evaluated for an experimental group of 48 salesmen whose earnings records were available for the full year of 1938, and who had filled out the application blank at the time of their employment. Six of the 31 items had to be thrown out because of the im-

SCORING WEIGHTS FOR PERSONAL HISTORY ITEMS

	Score
1. Age	
50	4
45–49	5
40–44	2
up–39	7
2. Height	
72″–up	7
70″–71.9	5
69″–69.9	4
up –68.9	3
3. Marital status	
Married	5
All others	3
4. No. of dependents	
4 or more	0
3	3
2	6
1	7
None	3
5. Thousands of Ins.	
10 or more	5
5 to 10	6
1 to 5	3
None	6
6. Amount of debts	
None	4
Current	6
$500 or more	5
7. Years of education	
Grades 1–8	6
9, 10, 11	3
12, Col. 1	6
Col. 2, 3	0
Col. 4, more	5
8. Number of clubs	
None	6
One	4
Two	6
Three, more	3

	Score
9. Years on last job	
Less than 1	5
1 to 1 yr. 11 mo.	1
2 to 2 yrs. 11 mo.	3
3 to 3 yrs. 11 mo.	6
4–5 yrs. 11 mo.	8
6–9 yrs. 11 mo.	10
10 or more	5
10. Experience in maintenance	
None	3
Any amount	6
11. Average No. Years on all previous jobs	
1 – 2½	3
3 – 6	5
6½–10	8
12. Average monthly earnings on last regular job	
Up to 150	5
150–199	4
200–249	8
250–349	1
350–399	5
400–up	6
13. Reason for leaving last regular job	
Still employed	10
Job discontinued (depression) (Co. folded) (also illness and circumstances beyond man's control)	7
To better self (positive reasons)	5
Was let go—dismissed (but if because of conflict with management, score as negative reason)	4
Negative reasons (friction)	2

CRITICAL SCORE = 62

The experience of the company is that 70% of those scoring above 62 are still working, while only 30% of those scoring below 62 are still employed.

possibility of objective scoring. The remaining 25 items were evaluated on the basis of differentiating between the upper and lower half of the sales force when 1938 earnings were used as a criterion. Twelve of these 25 items were found to be totally invalid on this basis. The 13 valid items remaining were then given scoring weights on the basis of the percentage of men giving each possible answer who fell within the limits of the upper half of the sales force in terms of 1938 earnings. For example, for the item of "Age" the scoring weights listed below mean that 70 per cent of the men who were under 40 years of age when employed were later in the upper half of the sales force in earnings, while only 20 per cent of those between the ages 40–44 were later above average in success. The scoring weights listed above are regarded as tentative, and as applying only to salesmen at Tremco Manufacturing Company.

Total scores on the 13 items obtained for each of the 48 salesmen in this experimental group correlated .67 with the 1938 earnings criterion. Critical total scores were then set to indicate a "danger zone" in employing new salesmen. In actual practice the sales managers did not adhere rigidly to these critical scores. This was in line with the established policy because of the recognized unreliabilities of scoring weights based on such a small sample.

During 1939 and 1940, 65 senior salesmen were employed. A follow-up study has just been made to determine what the effect of strict adherence to the critical scores would have been with this group. Since many of these men had been recently employed, it was possible to establish only a rough two-fold criterion—those who had actually failed and were no longer employed, and those who were still working. The results are as follows:

Of the 65, 33 were still working, 32 had failed.—Ratio 1 to 1.
If critical scores had been followed rigidly,
only 25 men would have been employed, of
whom 20 are still working, 5 have failed.—Ratio 4 to 1.
40 of the 65 would *not* have been employed.

13 of these are still working, 27 failed.—Ratio 1 to 2. Of these 13, however, 8 are regarded as probable failures by the end of the year.

These 13 items, together with some new experimental questions, constitute the new "Personal History Record" form. Critical scores on this form are now regarded as crystallizing the experience of the company and rendering it usable.

3. *Study of interviewing methods and rating forms.* At the present time the procedure for employing senior salesmen centers in the use of a series of interviews. Salesmen are employed by one of three district sales managers, and at some point in the territory to be filled. The employment procedure has been standardized and is based on the principle of successive hurdles, any one of which may eliminate an applicant. Even during the first few months of employment there are hurdles which are designed to eliminate the probable failures as early as possible. Emphasis is thus placed on sifting out the unfit at each step of the procedure rather than on "picking a winner" as the result of snap judgments in the early interviews. This tends to slow up the process of employing and to introduce checks and caution. Stated briefly, the procedure is as follows:

(1) Announcements of job vacancies state specific qualifications required, so that the obviously unfit will not apply.
(2) A telephone interview eliminates those who apply but do not meet minimum specifications.
(3) The first interview (about 45 minutes) serves to get specified objective data about the applicant's experience and history; to give him the facts about the job; to rate certain qualifications on a standard rating form.
(4) The new Personal History Record form is filled out by all men who are not eliminated by the first interview. The form is scored. Those falling below the critical score are eliminated.
(5) The second interview (about 2 hours) serves to answer any further questions the applicant may have about

the Tremco proposition; to elicit more subjective and personal information about the applicant's motivation, work habits, emotional background, character structure; and to form a definite basis for a more analytical rating of his qualifications on a standard rating form.

(6) Administration of a newly devised psychological test (The Tremco Test) to applicants still under consideration. Tentative norms and critical scores have been established.

(7) References are interviewed personally if possible. The applicant's reliability, sales record, and reasons for leaving are especially checked. Commercial credit reports are obtained on the final candidates. This step is not hurried, but when feasible the actual employment is delayed a week or two to permit both the applicant and the sales manager to "cool off."

(8) The third interview serves to give further specific information about the conditions of employment; and to secure any information about the applicant's qualifications concerning which there may still be any doubt. Employment usually takes place at this point.

(9) Employed men from various districts come to the home office for a training school. Ratings based on school performance are being accumulated for later validation against the earnings criterion. It is to be hoped that men who are likely to fail may be eliminated at this point.

(10) New men receive field training by a sales manager at intervals of a week or two. Detailed analyses of progress become the basis for further ratings of potentialities.

(11) A combination criterion of minimum earnings and number of sales per month, from the third to the sixth months of employment, serves as a final hurdle for the elimination of ultimate failures.

Since this procedure places considerable weight on interviews and ratings, an effort has been made to improve these methods. Although quantitative measures are being developed to evaluate certain of the qualifications, we are still dependent upon the clinical interview as a means of getting at the emotional and motivational aspects of sales aptitude.

The problem then was (a) to define the objectives of each interview—the qualities to be observed and rated, (b) to develop an interview procedure which would elicit the type of information upon which to judge these qualities, (c) to develop a form which would facilitate the analysis of interview data, and the recording of ratings, and (d) to instruct the sales managers in the use of these methods and forms.

The first of these problems was approached by setting up specifications for Tremco salesmen. Several lines of evidence were used to obtain this information. Successful and unsuccessful older salesmen were observed by the psychologist under actual working conditions in the field. Immediately after the dismissal of a salesman, an analysis was made by his sales manager of the specific causes of his failure. Conferences were held with the sales managers to obtain their judgment of important qualifications. An intensive clinical interview (by the psychologist) was arranged for all newly employed salesmen, as well as for many of the older salesmen. Pooling information from these sources has resulted in the following tentative list of qualifications to be rated in the interviews.

(1) Capacity for day-after-day plugging
(2) Money-mindedness, order-consciousness
(3) Capacity for social and emotional adjustment
(4) Capacity for intelligent planning of his work
(5) Practical maintenance sense
(6) Singleness of purpose in life, well-channeled motivation
(7) Financial position—need, but no financial tangles
(8) Physical energy and vitality
(9) Articulation—ability to think clearly and express his ideas simply and convincingly
(10) Honesty, reliability (to be judged from a commercial credit report, not from the interview).

An interview rating chart has been developed which defines each of these qualities, not in terms of the sales job, but in terms of specific information which is obtainable in the interview, and on which the rating is to be based. For example, the first item reads as follows:

(1) Capacity for day-after-day plugging

On previous jobs did he work regular hours; make 8–10 calls a day; dig up his own prospects; work on his own without much supervision?

Does he have courage to "stick it out"? Has he changed jobs frequently? Has he completed what he set out to do? Does the darker side of our proposition make him hesitate?

In the interview, was he business-like, direct, aggressive, decisive? Was he prompt? Did he have other things planned for the day?

As a boy, was he trained in regular work habits; given jobs and held responsible for them; required to earn his own money? Was home discipline firm, regular, dependable, but fair? Did he shift for himself; get his own jobs?

Such information is not likely to be secured unless the interview is carefully planned, and yet proceeds in a free and natural manner. In order to achieve this, a detailed outline for the second interview was developed. The outline contains 140 suggested questions or topics, and a method for leading so naturally from one to another that the impression of a casual conversation can be retained. It was neither hoped nor desired that all interviewers follow this outline in every detail in every interview. It has served more as a suggested procedure for getting from one area of discussion to another, and to insure the covering of all areas. The detailed outline need not be reproduced here, but its scope may be inferred to some extent by listing the principal areas of inquiry in order.

Work history, beginning with his last job and working back to the school years

Educational history, working back to elementary school and early childhood

Family background in childhood, emotional adjustment, and ego-structure

Work, money, and discipline in the home—going from the handling of money to the

Present economic, domestic, and motivational situation

Intensive study of one or two recent jobs, preferably those most closely related to the Tremco job

The Tremco proposition; his reaction to it; the sales manager's reaction to him as an applicant as a means of drawing him into an attempt to sell himself

A check list of items concerning his behavior during the interview, for the sales manager to go over immediately after the interview.

There are approximately 18 suggested leading questions under each one of these headings. The order of questioning is rather easy to remember because it is chronological and natural.

The principal distinguishing characteristic of this interview outline is its underlying assumption that the work habits and personality traits of an applicant can often be judged more accurately from an account of his early childhood than from his own carefully planned and glowing account of his recent job performance. Applicants expect to be questioned about their job experience, and so come prepared with a smooth set of rationalizations. But when they are led into a casual conversation about their boyhood activities and family relationships, they become pleasantly reminiscent and are less "on guard." The qualities which are being judged are also less evident to the applicant. It may be difficult to uncover the friction which caused his dismissal from his last job, but the roots of such a trait may be noticed in the way he tells about his domineering older brother or father.

While considerable thought has been given to the development of these interviewing and rating methods, there is at present only very indirect evidence of their effectiveness. Partial validation of the interview ratings has been obtained by checking the ratings at the time of employment against the sales manager's analysis of the causes of failure for a group of men who have been dismissed. Final validation will be made when a sufficient number of ratings on newly employed men have been accumulated to permit correlation against the criterion of later earnings.

4. *Study of psychological tests as predictors.* Some years ago the company tried out various so-called tests of intelligence, observation, attention, and imagination, on the sales force. In line with other investigations, no correlation was found between these tests and sales volume.

At present the Bernreuter Personality Inventory and Strong's Vocational Interest Blank are being administered to all salesmen at the time of employment. Statistical validation of these measures will be undertaken when a sufficiently large sample of such cases have established a job record. On a preliminary group of 30 cases the rough indications are that neither the "neurotic" nor "dominance" scores on the Bernreuter would have differentiated between those who later failed and those who are still working. It is possible however that such tests may prove useful in spotting an occasional extremely maladjusted individual and thus adding one more hurdle for the elimination of the unfit. We are continuing the accumulation of data on both of these tests.

The Tremco Test was devised to measure certain aspects of sales ability which could not be judged very well in the employment interviews. It contains three sections.

Part 1. How to handle difficult sales situations.
This section describes 22 of the most common sales objections and difficulties met by the Tremco salesman, together with multiple choice solutions.

Part 2. How to judge maintenance needs.
A series of verbal and pictorial descriptions of building maintenance problems, intended to measure not only his practical maintenance sense but also his judgment of the best approach to the sale.

Part 3. How to figure maintenance jobs.
A series of simple arithmetic calculations of the amount of material needed, and the cost of doing certain repair work.

The test is being used for the first time this year. It was listed as step 6 in the employment procedure described on

page 24. Preliminary item validation and tentative norms have resulted from administering it to a group of college men and a group of employed Tremco salesmen. Final validation must await the accumulation of enough test scores on newly employed men who have subsequently established a sales record.

It should be pointed out that two other problems contribute to the difficulties of developing an effective sales force, namely, the improvement of the supply of sales applicants, and the improvement of the program of training and supervision. The first of these is not an easy problem to solve when applicants with some maturity in age and sales experience are needed. A program for training junior salesmen has however been undertaken with this in mind. The second of these problems has received a good deal of attention at Tremco Manufacturing Company. Sales training by means of company schools and supervised field training has been coordinated as a part of the selection problem. It is hoped that training reports will become increasingly valid as predictors of the degree of ultimate success of salesmen.

PROCEDURES FOR THE SELECTION OF SALESMEN FOR A DETERGENT COMPANY

JAY L. OTIS
Western Reserve University

SELECTION procedures which can be used immediately to select better salesmen, and advice concerning the methods which should be followed in dealing with the present sales force were requested by the president of a small detergent company. The purpose of this study is to describe the methods which were established for the selection of salesmen for this company, and to describe the procedure for validation which is now in progress.

The company in which this study is being made is relatively small and sells soap and alkalies to laundries, metal cleaners to the metal industry, textile cleaners to textile plants, and heavy chemicals to the various processing firms. Each item is sold by a separate department. This study is concerned with the laundry department only. At present 24 salesmen cover the entire United States. Rarely are more than 5 or 6 salesmen hired a year for this department.

SALES CONFERENCE

A sales conference was called, and was attended by the sales executives of the various departments in the company. One day was spent on the problem of selection of laundry salesmen and opinions of each executive were classified, and those which were practical were set up as hiring standards. At that conference the job was defined, certain specifications were set up, and the duties of the salesmen were specified. Since two field supervisors had just been appointed, the job was studied quite thoroughly in this conference. Briefly outlined, the job is as follows:

Each salesman is given a territory and with one exception these territories follow state lines. The detergents are sold direct to laundries by the company salesmen and are also sold through jobbers or distributors. Credit for all sales in a territory is given to the laundry saleman assigned to that territory. Each salesman is required to work with the distributor salesmen and to train them in the selling of the product. In addition, since detergents are chemicals, and errors are made in their use in laundries, the salesman must be able to correct mistakes, and to demonstrate and install the product in the laundry washroom. The salesman's job can be divided into three parts:

1. Selling detergents to laundries
2. Training distributor salesmen in the use and sale of detergents
3. Servicing and demonstrating the product.

This third function involves superior knowledge of washroom practice; hence laundry experience is usually necessary.

ESTABLISHING PRELIMINARY HIRING QUALIFICATIONS

The following selection items were chosen from a list submitted by each sales executive present at the sales conference:

1. Hard worker (capacity for work)
2. Personality (appearance, size, manner, etc.)
3. Sales ability (proven and obvious, experienced men only)
4. Knowledge of laundry business and washroom practice
5. Teaching ability (training jobber salesmen)
6. Business knowledge (jobber problems, contracts, credit, etc.)
7. Acquaintance with territory
8. Mental ability
9. Physical vitality (health)
10. Speech (voice important).

Each of the above ten points is considered in hiring each new salesman. Preliminary techniques have been drawn up and are being validated by studying their usefulness in the hiring situation.

PRELIMINARY HIRING PROCEDURE

Experience has shown that it is extremely difficult to find trained and qualified sales personnel in the detergent field. To solve the immediate problem of selection, the preliminary hiring procedure was divided into five main parts. It can be seen that an attempt was made to include all of the minimum qualifications listed above and for certain of these qualifications the employment procedures secure the necessary information in several ways. The procedures involve the use of an application blank, an interview blank, an interview rating scale, reference form, and psychological tests.

The application blank is designed to serve two purposes. First, it is very comprehensive and should furnish valuable information about the applicant. Second, certain items on the blank have been assigned scoring weights and preliminary results indicate they have some value in prediction.

The applicant interview form was designed to describe certain fields or areas of investigation and it possesses the characteristics of the free interview and some of the desirable qualities of the standardized interview. Since the form was designed to offer interview suggestions only, no interviewer is required to cover all of the minor items. He is required to cover the major headings. The Applicant Interview Form is shown as Form I.

FORM I

APPLICANT INTERVIEW FORM

Name of Applicant..... Date.....

Address ..... Interviewed by.....

FAMILY BACKGROUND

Are parents still living? Father's occupation? Mother's occupation? Father's educational background? Mother's? What sort of person was father? Does applicant depend upon him?

Are there brothers and sisters? Younger or older? Occupations of brothers and sisters? How successful is each? Was there competitive stimulus within the family and are there any signs of extreme dependence?

Regular chores around the house as a boy? Taught work habits and work attitudes? How get spending money? When start savings account?

Learn value of the dollar in home? In general, was family background such that it is an asset or liability to applicant?

EDUCATIONAL HISTORY

Which subjects liked best? Which disliked most? In general, preference for science and math or English and social studies? Interested in practical or theoretical or artistic things?

Extra-curricular activities. Sports, Dramatics, Fraternity, Social. Offices held. Work while in High School? In general, an average student, below or above average?

Major and Minor in College. (Same questions as above.)

Taken any special courses since leaving school? Speech. Selling. Business subjects. What did applicant get out of them? Any courses in mind applicant would like to take?

PRESENT DOMESTIC AND ECONOMIC SITUATION

Married? Children? Any other dependents? Age, sex, schooling of children? Wife working? Any enthusiasm in talking about family?

What is wife's attitude toward (changing job) proposition? What is wife's attitude about being away from home? Will wife move to new territory willingly? Are wife's ideas important? Dependent on her?

Own home or rent? Mortgage? Current bills at stores? (Credit.) Any other indebtedness? Life Insurance? Loans against it? Savings account now? Any reserve? Any other sources of income? Will living standard need to be adjusted?

What are principal hobbies or interests? Is life centered around the trivial such as sports, good times—or around getting ahead, building a future? Belong to many clubs? (Not too many scattered interests better.)

Does applicant appear emotionally upset for any reason? Health? Applicant has a home, family, income, etc., what else does he want? What working toward?

WORK HISTORY

Begin with first full time job and work up to present application. Any jobs related to detergent selling? Any laundry? Any metal? Heavy Chemicals? Maintenance? Textile? Food and Beverage?

What type customer called on? Any distributor contacts? Work with distributor salesmen? Demonstration? Service? Any training experience?

How many jobs? How long on each job? Reasons for leaving each job (Friction, Dismissals, Good Judgment)? What job liked best? (Interest and enthusiasm about jobs?) Any strong loyalty or attachment to any employer? Has work experience been of value in job applied for? Approximate earnings on each job? (Past peak?)

Is he ambitious? Wide shifts in type of work? Can applicant handle technical aspects of job without training? How much training necessary?

INTENSIVE STUDY OF ONE OR TWO RECENT JOBS

Just what type of product was or is he handling? Does applicant know his company and his line thoroughly?

Describe a typical day's work. Just what did he do? How did he find his prospects? How did he make new contacts? What type sales approach did he use? ("Canned" talk, plan each approach, or trust to luck?) About how many calls a day? How long a day? Work regular hours—early and late?

How heavy were expenses? Keep any personal record system? (Systematic.) Has applicant testimonials or records of sales or earnings? Was story told enthusiastically? Did he stress money earned or sales made, rather than number of friends or quality of products?

REACTION TO PROPOSITION

Does applicant have any questions to ask about our proposition? (Note type of question.) What is general reaction to our proposition? What appeals particularly about it? (Making money. Quality product. Company reputation. Advancement opportunities. Future security.)

What is likely to be chief difficulty or handicap? (Is thinking hard-headed, planful, sound, wishful, emotional, over-enthusiastic, confused?)

NOTE ALSO

Physical appearance. Vitality and energy. Business-like interview. Well-organized and clear speech. Any obnoxious manners? (Egotism, indecision, self-critical, apologetic?)

What is applicant's attitude toward work, associates and superiors? Do hobbies and recreational activities, religious and political beliefs, closest friendship and social contacts indicate rounded, well balanced life?

Will applicant be able to deal with customers? Can he handle distributor contacts? What will his effect be on distributor salesmen?

FORM II

RATING CHART FOR APPLICANTS

1. KNOWLEDGE OF BUSINESS:

The applicant must know the laundry, metal, maintenance, textile, etc., business to sell products. If considerable training will be required, rate him low. If he can sell without further training, rate him high. Keep demonstrations, check up service and trouble shooting in mind when judging. Check your judgment below.

Excellent ☐	Good ☐	Average ☐	Poor ☐	Very poor ☐
No Training Required		Some Knowledge		Little Practical Knowledge

2. MENTAL ABILITY:

Mental ability does not necessarily mean educational level. Will applicant be able to learn to sell our products? Is he bright and alert? Has he imagination? Will he show initiative and be helpful to the distributor? Will he be able to plan intelligently? Check your judgment below.

Excellent ☐	Good ☐	Average ☐	Poor ☐	Very poor ☐
Very Superior	Superior	Ordinary	Slow	Dull

3. BUSINESS KNOWLEDGE:

Is applicant a business man? Does he comprehend the problems in development, purchasing, manufacturing, accounting and sales in a corporate business? Does he understand that our expense must be within our income, that there must be profit for stockholders? Does he possess enough business background and knowledge to represent us to distributors? Check your judgment below.

Excellent ☐ Good ☐ Average ☐ Poor ☐ Very poor ☐
Considerable Business Knowledge — Little Business Knowledge

4. PERSONALITY:

Consider how applicant impresses others by his manner, bearing and tact. Does he smile readily? Does he talk too much or too little? Is he blunt? Too much self-importance, ego? How will he impress his customers? Will he work well with distributor salesmen? Will his personality be an asset in selling? Check your judgment below.

Excellent ☐ Good ☐ Average ☐ Poor ☐ Very poor ☐
Outstanding Personality — Not Outstanding — Personality Handicap

5. SPEECH:

Does applicant express his ideas clearly and easily? Speech need not be polished or grammatically perfect, but thoughts should be clearly and easily understood. Is he interesting? Enthusiastic about what he says? Speak to the point without wandering? Voice should be satisfactory enough not to detract from ideas. Check your judgment below.

Excellent ☐ Good ☐ Average ☐ Poor ☐ Very poor ☐
Clear Expression — Satisfactory — Questionable Speech

6. SALES ABILITY:

We are interested in sales ability as it affects the sale of our products. Has applicant convinced you that he has successfully sold for someone else in a difficult sales situation? Does he have a "nose for business"? Will he need help in difficult situations? Is he self-reliant? Is his best selling ahead of him or behind him? Does he have records or proof of his ability to sell? Check your judgment below.

Excellent ☐ Good ☐ Average ☐ Poor ☐ Very poor ☐
Outstanding Sales Ability — Average Sales Ability — Questionable Sales Ability

7. TEACHING ABILITY:

A salesman must work with distributor salesmen. Has applicant ever demonstrated his ability to train? Was he successful? Can he present ideas clearly? Does he have the necessary teaching ability to train distributor salesmen to sell our products? Check your judgment below.

Excellent ☐ Good ☐ Average ☐ Poor ☐ Very poor ☐
Oustanding Training Ability — Training Ability — Questionable Training Ability

8. INDUSTRY:

Capacity for day-after-day plugging. On previous jobs did applicant work regular hours; make sufficient calls a day; dig up own prospects; work without supervision? Do you believe that he will be a hard and industrious worker? Check your judgment below.

Excellent ☐ Good ☐ Average ☐ Poor ☐ Very poor ☐
Hard Worker — Work When Pushed — Questionable Worker

9. PHYSICAL VITALITY:

Does the applicant impress you as having a surplus of physical energy? Is he vigorous, active, not sluggish? Is he physically large and impressive? If he is small in height or weight, he should make up for it in energy and health. Will he be able to travel, work hard? Check your judgment below.

Excellent ☐	Good ☐	Average ☐	Poor ☐	Very poor ☐
Strong	Energetic	Probably Satisfactory	Slow	Sluggish

10. ACQUAINTANCE WITH TERRITORY:

Does applicant know the territory for which he is being interviewed? Does he know customers in any territory? If not acquainted with prospects, rate low. If knowledge of territory and prospects is good, rate him high. Check your judgment below.

Excellent ☐	Good ☐	Average ☐	Poor ☐	Very poor ☐
Well Acquainted with Territory				Little Knowledge of Territory

11. SUMMARY JUDGMENT:

In making a final judgment about this man, keep in mind the situation which exists at the present time. You should know why the man is a good prospect as well as why you believe the man to be a poor prospect. Record your judgment and reasons below.

Excellent ☐	Good ☐	Average ☐	Poor ☐	Very poor ☐
Unqualified recommendation		Qualified recommendation		Not recommended

Reasons for Summary Judgment.....

.....

.....

.....

.....

.....

The rating chart, Form II, is found on the reverse side of the interview form and consists of those items which were mentioned as being important during the sales conference. The items are defined, and a space is provided for the recording of judgment. This rating chart gives the interviewer an opportunity to make a series of independent judgments concerning the qualifications of the applicant. Material from the application blank and the interview data are used as the basis of the ratings.

The eleventh item on the rating scale, The Summary Judgment, gives the interviewer an opportunity to decide whether

or not the man should be employed. A space is provided on this rating scale for the reasons for the summary judgment. Applicants are interviewed both in the field and in the home office, and it is essential that the reasons for acceptance and rejection be known so that the person who makes the final decision will be able to evaluate the ratings.

The choice of references for sales positions includes educational institutions, former employers, former customers, and business acquaintances. A reference form has been constructed for use with employer, business, and customer references. There is a reference letter on one side of the page and a rating chart on the other side. The rating chart asks rather specific questions and includes material on the following areas: industry or work habits, business knowledge of applicant, personality characteristics, character qualifications, teaching ability, and sales ability.

The field supervisors are instructed to make personal visits to former employers, business acquaintances, and former customers of applicants whenever possible. Telephone calls to references are also used. In cases where the personal visits and telephone calls are used, the rating chart of the reference form is filled out for the applicant for consideration in the home office.

Each applicant is given the Strong Vocational Interest Blank, the Bernreuter Personality Inventory, and the Bureau Test VI. The tests are scored and interpreted by the consulting psychologist, and the test interpretation forms a part of the preliminary hiring procedure.

A final selection form has been constructed and all of the results of the five steps, outlined and discussed above, are recorded on this blank. The application blank score, the results of each interview as shown by the interview rating scale, the references, and the psychological test interpretation are summarized. The applicant is either recommended for employment, or he is questioned and is hired, if a more suitable applicant cannot be found, or he is not recommended for employment.

THE EXPERIMENT

An attempt is being made to standardize the selection procedure so that more precise prediction can be obtained than is possible with the above described preliminary selection technique. Valid techniques require the standardization of criterion variates and data are being collected which not only are of value for the criterion but are also of value in training and in making suggestions which better the job performance of the present personnel.

THE CRITERION

No two sales territories are alike. One salesman, for example, works in New York City, and another salesman sells in three large western states. Competition varies, and the distributors and jobbers associated with the company introduce a variable which makes it difficult to measure sales performance.

Both objective and subjective criteria are being used; however, more attention is being paid to the objective measures. The present accounting system makes it extremely difficult to obtain an exact record of the sales made by any one salesman. Since all distributor sales are credited to the salesman in the territory where the detergents were sold, and since all the direct sales made by the laundry salesman are credited to the distributor who fills the order, the credit for any one sale or for the sales for a particular period of time is difficult if not impossible to determine. The number and quality of distributors varies and a poor salesman with good distributors may have a good record, while a good salesman with few or poor distributors may have a poor record.

In spite of these difficulties, criterion data including such items as gross sales made in each sales territory, profit made in each territory, sales cost, and number of new accounts sold have been collected and are being studied in respect to territory differences. Preliminary criterion variates, such as gross sales and selling cost, have been established for each salesman. These data are highly reliable.

Interest in job performance influenced the company to institute a more detailed daily sales report and nine months' data are now available for the present salesmen. The data which are being collected include total number of calls, average calls per day, miles per call, description of the accounts called on, number of calls with jobber salesmen, and the amount of time in hours spent on service in the laundry washroom. Reliability coefficients are all above .78.

When the job performance record is compared with the sales performance, relationships and lack of relationships are found which are important. There is, for example, no relationship between average number of calls per day and gross sales, selling cost, and profit. Data show that the men who open the most new accounts make on the average fewer calls per day. This indicates that the men who do the best selling job either take more time per call or spend more time planning their calls. Since salesmen are required to spend time servicing laundries, it is of interest to compare the service time with other performances. There is a fairly high relationship between service time and the sales variates. For example, a correlation of .51 was obtained between service time and profit in the territory. The salesmen who spend the most time in service calls tend to be better on the sales criterion variates.

A series of ratings were obtained on each salesman. The relationship between this subjective criterion and the job performance variates is positive but not significant.

VALIDITY

Many studies have shown the value of application blank or personal data items in the selection of salesmen. Such items as telephone, marital status, number of children, number of dependents, living conditions, education, and special courses taken were assigned weights, selected from weights obtained by other investigators, and a personal data score was obtained for each employed salesman. Validity coefficients of .31 and .37 were obtained when the personal data scores were correlated with selling cost and gross sales.

No relationship was observed when the Bernreuter Person-

ality Inventory scores were correlated with the gross sales and selling cost variates. The obtained scores indicated a high degree of homogeneity, and this might account for the absence of relationship.

There was a positive correlation, .50, between the total of the Life Insurance Sales and Real Estate Sales items on the Strong Vocational Interest Blank and selling cost. Further study will be needed to establish the validity of all of the psychological test items, since data are available for only 17 cases at the present time.

Because of the relatively few men hired each year, no attempt will be made to combine the various criterion variates. A separate predictor battery will be standardized for each predictable variate. Final validity will be determined by the follow-up procedure.

SUMMARY

In summary the work which has been done in the selection of salesmen for large companies is well known. In the small company, employing few salesmen, a mistake in selecting one man can spoil a large territory so that a relatively large part of the total income to the firm will be lost for a period of time. Since this is true, techniques must be found which will do a better job of selecting for small companies than the job which large organizations are now doing.

This attempt involves an intensive study of each applicant as outlined in the preliminary selection procedure. In addition, test batteries designed to predict the various criteria such as sales, selling cost, and profit, and the various job performance variates such as number of calls per day and service time are being standardized. Preliminary results indicate that the validity of these batteries on the employed group is satisfactory. Provision has been made for adequate follow up.

This procedure is an attempt to substitute time and extensive data for a large experimental sample. The use of independent criteria and a test battery for each criterion rather than a composite criterion with a single test battery should increase the validity of the selection procedure.

SELECTION RESEARCH IN A SALES ORGANIZATION

THOMAS M. STOKES
Metropolitan Life Insurance Company

IN considering what I could say to your organization about selection research I immediately eliminated any discussion of the research methods followed or the research results secured in my Company. Dr. R. S. Schultz who has supervised this research is better qualified to explain it to you and tell you how it resembles and how it differs from similar research projects carried on in other organizations. I am interested but not actively engaged in Selection Research, and I do not propose to plunge—at least intentionally—into psychological water that is over my head.

What I shall attempt to do will be to touch on a few problems of selection research in a sales organization as they appear to an individual who was trained over twenty years ago in psychology and in the scientific approach to human problems, who was personally active in selection research during and following the first World War, who has retained some of the interests and attitudes of the psychologist, but who for some sixteen years has been primarily interested in improving the training of sales personnel. These problems fall under four headings, namely:

1. The significance of the job environment—with particular reference to training—in selection research;
2. The need for dynamic research in selection;
3. The cooperative nature of selection research;
4. The practical application of selection research.

THE SIGNIFICANCE OF THE JOB ENVIRONMENT—WITH PARTICULAR REFERENCE TO TRAINING—IN SELECTION RESEARCH

Selection in a sales organization is the art of (1) determining whether an individual has certain abilities, interests, and behavior patterns which make his success in *a new job environment* reasonably certain; and (2) identifying an individual whose abilities, interests, and behavior patterns make his success in a new job environment more probable than that of other individuals being considered for the same job. Both kinds of selections are made in sales organizations.

In arriving at objectives in selection research, we should consider not only the kind of work that is to be done but also the job environment. By the job environment of a salesman I mean those factors *which together with* the salesman's personal traits determine his length of service and the quantity and quality of his accomplishments. For example, the job environment of an Agent who works in a section where there is widespread unemployment naturally differs from that of an Agent in a community where the factories are working overtime. And, other things being equal, since the job environments differ, the criteria of success vary in the two cases.

Now training may be a powerful factor in the job environment. Training in a sales organization is the art of helping a salesman learn, that is helping him change his behavior in such a way that he can be more successful in his job. Training, as we think of it, results in the salesman acquiring certain attitudes, habits, skills, and knowledge which result in a different pattern of behavior—a pattern which enables him to accomplish more in a given time and at the same time derive greater satisfaction from his work.[1]

In selecting a salesman for an office where there is no efficient training, we are selecting a man who will probably

[1] A more complete explanation of the writer's conception of training is published in two articles: "What Training Means," *Managers Magazine*, Nov.–Dec., 1937, and "The Training Process," *ibid.*, Jan.–Feb., 1938.

succeed because of qualifications he has prior to selection. But in selecting a salesman for an office where there is efficient training, we are selecting a man who will probably succeed not only because of his present qualifications but because of the training he will receive. In other words, we select men *for training in a particular job environment.*

Perhaps if I briefly summarize the results of training in increased earnings of a few Agents, I can better explain the fact that we feel we have definitely demonstrated in our Company, namely that training can be a tremendously powerful factor in the Agent's job environment. Each of the Agents that I shall refer to had been trained prior to the special training mentioned. This special training was conducted along lines that have been developed in our Field Training Division during the past eight years.

Agent E had been an Agent for three years when he received this special training. His average weekly income for a year and a quarter following this training showed an increase of 32 per cent.

Agent B received this training four years after his appointment as Agent. His average weekly income for one year following this training showed an increase of 56 per cent.

Agent F's special training occurred nearly four years after his appointment as Agent. The increase in his average weekly income for the next year and a half was 31 per cent.

Agent L received special training fifteen years after appointment. At the time he received this training he was below his District's average in Ordinary production. Following this training, although the Ordinary production of the District fell off, Agent L's Ordinary production increased in dramatic fashion, more than doubling in about eighteen months' time.

Increase in income and in sales are only two factors of improvement, but time does not permit a discussion of others.

Now I have introduced the subject of training because selection and training are two sides of a coin, and research in one should be accompanied by research in the other. The effective-

ness of training, by which I mean all managerial activities which help salesmen attain greater success, is a factor that must be taken into consideration in selection research.

THE NEED FOR DYNAMIC RESEARCH IN SELECTION

Selection research may be either static or dynamic. Dynamic research assumes a varying job. It not only recognizes that the jobs of different salesmen differ because of differences in job environments but that jobs and job environments change with the passing of time. These changes in the salesman's job like the changes in automobile models may be slight from one year to the next but are much more pronounced over a period of time. Just to mention one change since 1939 in our Agency organization, Agents are now being trained to sell an amount of insurance which when combined with benefits to a widow from Social Security will provide an income of a certain size. This means that due to a social change in American life, Agents in interviewing many prospects are now using a new kind of sales presentation.

Static research more or less assumes standardized jobs and that once a selection procedure has been evolved, the research has been completed. Dynamic research implies that while the sales organization has a better program of selection than formerly, the research man has the responsibility of improving it and keeping it effective. If our selection research is to be vital and of the greatest practical value, it should be more dynamic in the future than most of it has been in the past.

Because of the variations and changes in job environments and criteria of success, the problem of subjective ratings by Managers should receive considerable attention from the psychologist if the initial research is to be properly followed up. In our research, we used three criteria that could be expressed quantitatively, namely, length of service, amount of Ordinary and Accident and Health insurance placed, and increase in Industrial insurance sold. These quantitative criteria, however, do not accurately measure the success of the salesman.

Other factors such as the *quality* of the Agent's service to prospects and policyholders and his cooperation with management and fellow Agents are important.

The next phase of our research which we are just beginning is to compare an Agent's selection test scores and other selection forms with objective criteria *plus* subjective managerial ratings after Agents have worked for a period of three months and longer. We are familiar with the criticisms that have been made against subjective ratings; but until we can satisfactorily express qualitative criteria in objectively measured terms, we expect to continue to use and to endeavor to improve these personal ratings.

THE COOPERATIVE NATURE OF SELECTION RESEARCH

The approach of the psychologist should be to study the personnel problems in a sales organization and determine what needs to be done in order to improve the selection of salesmen. Having determined the organization's needs, the psychologist can then consider whether one or more psychological techniques can be used in helping management solve its problems. I am convinced that in some sales organizations selection research will be sterile until other management difficulties—especially training difficulties—have been taken care of.

The purpose of selection research is not to develop good selection methods to replace bad ones. It should be to help the sales executive improve his present methods which may already be far superior to those of many other organizations. Before the psychologist enters into the picture, the sales executive may be weeding out physical unfits by medical examinations. He may be weeding out moral unfits through securing inspection reports from an outside organization. It is quite possible, as in my Company, that statistical studies of the application blanks of salesmen and their length of service and production may have been made, and that an intelligent review of application blanks may be improving the selection of salesmen. Some Managers may be using methods of re-

cruiting and interviewing applicants which enable them to select better than average salesmen.

The purpose of selection research is to build on present methods, improving them as needed and adding the new selection tools and methods that will result in still better selection. My Company is now using the Selection Procedure which Dr. Schultz and the Psychological Corporation have helped us develop.[2] We are convinced that it is definitely helping us select better Agents. But whether progress has been made has been largely due to the fact that the present methods and tools have been developed cooperatively by psychologists and members of our own organization.

Executives in my Company have devoted time and thought to the selection problem and to the research progress made over a period of years by the Psychological Corporation. A number of our Managers helped us by giving preliminary tests to applicants. These Managers knew that the tests were experimental but they shared our interest in developing better methods.

Another example of the cooperative nature of our research was the participation of four of our Instructors in developing interview procedures and a Calculation Test. These Instructors—none of whom were psychologists—spent about 300 hours on this work over a period of eight months, under the guidance of Dr. Schultz.

THE PRACTICAL APPLICATION OF SELECTION RESEARCH

One of our objectives was to develop a selection procedure which would not require a psychologist to operate. We endeavored to use the language of the layman, not of the technician. Nevertheless, we realized that to increase the probability of efficient use by Managers, three things were necessary:

[2] *Cf.* R. S. Schultz, "Test Selected Salesmen Are Successful," *Person J.*, 14, 1935, 139–142, and R. S. Schultz, "Standardized Tests and Statistical Procedures in Selection of Life Insurance Sales Personnel," *J. Appl. Psychol.*, 20, 1936, 553–566.

(1) To show the need for the new procedure and to create interest in it.
(2) To help Managers acquire skill in its use by explaining it to them, demonstrating its use and observing them in its use.
(3) To help them retain skill in its use.

This last training activity is particularly important because of the low turnover in the Metropolitan Field Force. Many Managers have not had occasion to select a new Agent in the past year and some have not for a number of years. Lack of practice in using this procedure naturally results in loss of skill and necessitates retraining in its use.

We are in a favorable position to provide training and retraining in the use of our selection procedure because we have groups of Field Instructors working in all sections of the United States and Canada where there are Metropolitan offices. These Instructors not only receive special training when assigned to this work but receive additional training each year in all phases of their work which of course includes training men to use the Selection Procedure.

EXTENSION OF THE NORMS OF THE COLUMBIA VOCABULARY TEST[1]

GEORGE K. BENNETT AND SYDNEY ROSLOW
The Psychological Corporation

THE Columbia Vocabulary Test was constructed and standardized for grades 3 to 8 by Irene Gansl and Henry Garrett.[2] It consists of 100 sentence type multiple choice items in each of which the subject is required to select the response word most nearly synonymous with the designated stimulus word. Students in the latter half of the eighth grade obtain a mean score of 54 with a standard deviation of 15.[3] To explore the feasibility of extending the standardization of the test through secondary school, a group of recent high school graduates with superior academic records was tested. These 1212 graduates earned a mean of 74 with a standard deviation of 12. This seemed to us to justify the form for high school use.

Since the original standardization was based on pupils in New York City elementary schools, it appeared desirable to utilize New York City high school students for the extension of the norms. With the cooperation of The Bureau of Reference, Research and Statistics of the New York City Board of Education, Samuel Tilden High School was chosen as representative of the city's high school population.[4]

[1] Published by the Psychological Corporation, New York City.

[2] Irene Gansl, "Vocabulary; Its Measurement and Growth," *Arch. Psychol.*, #236, 1938.

[3] Irene Gansl, *op. cit.*

[4] The authors acknowledge with thanks the cooperation of Dr. Herbert Landry, Bureau of Reference, Research and Statistics, Dr. Abraham Lefkowitz, Principal, and Miss Olga Hingsburg, Guidance Counselor, Samuel Tilden High School and the many teachers who aided in administering the examination.

Through the assistance of the school administration and faculty, this test was administered simultaneously to all of the pupils (5101) present on March 4, 1940. These students were divided among the eight half grades and the two course options as shown in Table 1.

TABLE 1

Grade	General Course	Commercial Course
	N	N
9A	203	189
9B	271	206
10A	394	332
10B	367	333
11A	473	346
11B	340	263
12A	389	301
12B	423	271
Total	2860	2241

Each of these tests was then scored by the IBM test scoring machine and verified by hand according to the formula, rights minus one-quarter wrongs. Table 2 contains the means and standard deviations by grades for each course.

TABLE 2

Grade	General Course		Commercial Course	
	Mean	S.D.	Mean	S.D.
9A	55.5	15.2	52.4	15.9
9B	57.9	13.2	49.1	11.7
10A	66.2	12.2	53.9	13.0
10B	67.9	12.0	57.0	12.7
11A	72.0	11.9	59.5	12.8
11B	71.2	10.0	62.2	12.6
12A	75.2	10.1	67.2	11.8
12B	77.7	9.1	68.9	11.0

The fact that the mean score of the highest grade, 12B general, is approximately 2.5 standard deviations below the pos-

sible maximum appears to indicate that the test is suitable for use throughout high school. Skewness, according to the formula $\frac{3\,(M-Md)}{\sigma}$,[5] is $-.28$, indicating a slight degree of skew-

TABLE 3

The Mean Score for Each Age-Grade Combination

The upper figure is the mean score to the nearest whole point, the figure in parentheses below is the number of cases upon which the mean is based. Age is taken as that of the nearest birthday.

	Grade							
Age	9A	9B	10A	10B	11A	11B	12A	12B
				General Course				
18						67	69	74
						(28)	(44)	(103)
17				62	67	69	74	79
				(23)	(70)	(132)	(136)	(177)
16			60	66	71	73	77	80
			(71)	(129)	(209)	(141)	(157)	(112)
15		56	66	69	75	73	80	
		(128)	(199)	(158)	(151)	(37)	(49)	
14	56	60	71	73	83			
	(154)	(108)	(102)	(52)	(22)			
13	60							
	(32)							
				Commercial Course				
18						52	61	68
						(25)	(50)	(78)
17				54	56	62	67	69
				(26)	(91)	(122)	(153)	(136)
16		40	50	54	60	66	72	74
		(40)	(73)	(153)	(176)	(89)	(86)	(42)
15	53	50	55	61	66	72		
	(39)	(95)	(195)	(125)	(54)	(19)		
14	53	52	59	64				
	(119)	(63)	(45)	(18)				
13	53							
	(24)							

[5] J. P. Guilford, *Psychometric Methods*. McGraw-Hill, New York, 1936, p. 104.

ness. In view of these facts the decrease in standard deviation with advancing grade probably indicates a tendency toward greater homogeneity in respect to vocabulary in the later years of high school.

Each student indicated the date of his birth, permitting comparison on the basis of age. However, since high school students cannot be regarded as unselected representatives of their age groups, age means are of dubious value. Due to the large number of cases a more significant presentation (by age and grade) can be made. This is shown in Table 3 for both general and commercial students.

It will be seen that there are two reasonably regular progressions. For both general and commercial students, with age constant, vocabulary increases with grade and, with grade constant, vocabulary decreases with age. Both of these trends are consistent with the theory of the relationship between intelligence and academic status.

PSYCHOLOGICAL APPLICATIONS TO THE NEW FIELD OF TRAFFIC ENGINEERING

T. W. FORBES

Bureau for Street Traffic Research, Yale University

A VERY large proportion of the previous work involving psychological applications to automotive traffic and to traffic safety has involved driver testing or selection of drivers by various methods. The most familiar and outstanding objects of this work have been:

1. A selection of accident prone drivers so that they may in some fashion be prevented from endangering the lives of others.
2. The selection of drivers for commercial fleets, or the reduction of the accident records of these fleets.
3. The much discussed educational use of the driver test and interview technique.

Having been engaged in the driver test field previously, the writer is sympathetic with its aims and recognizes the importance of these applications. However, certain of these applications, of necessity, have a somewhat negative aspect, since their success means the elimination of certain drivers. This negative aspect is undoubtedly necessary for certain purposes, but where it is possible to employ them, the more positive approaches would seem to be preferable. The psychological applications to traffic engineering to be considered in this paper attempt to contribute to the positive side of the solution of some of these problems, and are felt to be highly promising.

A recent report[1] gives the results of a study of 25,000 drivers from the State of Connecticut carried out by the Highway Research Board under the direction of Dr. H. M. Johnson. A

[1] Secretary of Agriculture. Motor Vehicle Traffic Conditions in the United States. The Accident Prone Driver. House Document #462, Part 6, Govt. Printg. Office, Washington, D. C., 1938.

subsequent analysis of the figures from this study by the present writer shows that if a truly "accident prone" driver is defined to be one who has accidents in both the first and last three years of a six year period, only 1.33 per cent of these drivers were accident prone. This means that the vast majority of drivers (whom we will call "normal" drivers for convenience) sustain either no accidents or very sporadic ones. Certainly it would hardly be practical to regard a person who has one reportable accident and only one in six years as being worthy of being ruled off the road as a menace. The records show that too many people have one such accident and may never have another. This vast majority of the driving population who are by and large pretty good drivers must then be treated by other methods under the enforcement, educational and engineering organizations.

The enforcement officials and educators have the problem of modifying the behavior of the general run of drivers under present conditions, whether these conditions be good or poor. The educator also attempts to inculcate driving habits and states of mind looking toward better driving. The traffic engineer's position, however, is unique in that he has the opportunity of providing new traffic facilities better adapted to the motorists' needs and abilities. For this purpose the traffic engineer needs some psychological or "human engineering" data to parallel the data which are available to him on physical layouts and structures.

The psychological applications to traffic engineering discussed below represent an attempt to build up such a body of data for the practical guidance of the traffic engineer in furnishing facilities better suited to the needs of the motorist. The field of traffic engineering is a fairly new one, and is therefore somewhat unfamiliar to a great many people. It consists, in brief, of developing and applying engineering methods for the handling of traffic in order to make both city and highway traffic more efficient and at the same time more safe. Traffic engineers are being employed in both cities and states to

greater and greater extent. The more usual of the duties which are assigned to the traffic engineer's department are those of designing and erecting signs, timing traffic signals for the best flow of traffic, overseeing the marking of pavements for the routing of traffic, the redesigning of intersections to improve the safety and efficiency of the intersections, and developing new street and highway designs to obtain better operating characteristics. Under the latter heading fall the geometrical design of ramps, traffic circles, curves, bridge abutments, safety islands and other similar operating features. In the past, the traffic engineer has been faced with the task of carrying on all these functions with the physical data which are available and by means of the best guesses which he can make as to the human reactions which will be involved. He has done a surprisingly good job in most instances, but the leading engineers in the field have recognized the desirability of having actual factual data on the human aspect of the problems involved in a good, practical, usable form. The program of studies which is being described has been intended first, to devise and demonstrate methods; and second, to obtain practical results which may be of immediate interest and application for the traffic engineer in the field.

Two major questions are of importance to the traffic engineer. The first question is "What *can* the ordinary driver do?" and the second is "What *will* or what *does* the ordinary driver do?" Thus we have the two general fields of first, the limits of ability of the "normal" driver; and second, ordinary driver behavior under given conditions.

Certain studies have been undertaken which deal with the limits of ordinary driver ability.

1. *Studies of Sign Legibility.* Under this heading might be classed a recent study[2] done by our Bureau jointly with the Pennsylvania Turnpike Commission on the legibility of stand-

[2] T. W. Forbes and R. S. Holmes, "Legibility Distances of Highway Destination Signs in Relation to Letter Height, Letter Width and Reflectorization," *Proc. H'way Res. Bd.*, 1939, 19, 321–335.

ard destination sign letters of various heights and widths. The practical man who deals with signs has for some time wanted to know how to obtain a reliable measure of the distance the general public can see various types of signs. A series of test signs using letters of two different widths and of heights varying from 6″ to 24″ were devised. A technique developed earlier[3] was used. Groups of observers approached these from a distance and recorded the point at which they were able to read the sign, thus giving a distribution of legibility distances for each height and width of letter. Observations were carried out under both day and night conditions so as to give results for reflectorized and unreflectorized letters. To summarize, the study showed the need for at least 10″ narrow or 8″ wide letters in destination signs if placed 400 feet ahead of the hazard, and the need for 18″ wide letters if placed at the hazard. This size of letter was indicated by the distance at which the signs were legible, and by assuming that the ordinary driver needs at least 10 seconds warning at a 50 mile per hour design speed.

This assumption of 10 seconds is based on time for reading the sign, perception-reaction time, and a sufficient time for the slowing of the vehicle. On the matter of reading time, a previous study by Lauer[4] has indicated that for some signs this may take as much as 4 to 11 seconds. However, we have been able to demonstrate[3] with a shutter tachistoscope that a properly designed sign with minimum amount of copy may be seen in very much shorter time than this. In rough fashion it has been demonstrated that a 3-word sign can be seen in a single glance which consumes from 0.6 to 1.0 second. This figure is based on eye movement away from the road, fixation, and return to the road. More data are needed on certain of these problems.

[3] T. W. Forbes, "A Method for Analysis of the Effectiveness of Highway Signs," *J. Appl. Psychol.*, 1939, 23, 669–684.

[4] A. R. Lauer, "Improvement in Highway Safety," *Proc. H'way Res. Bd.*, 1932, 12, 389–401.

2. *Study of Arrow Legibility.* At the present time there is a further study of legibility of destination sign arrows in progress jointly by our Bureau and the Connecticut State Highway Department. Many other visibility factors applying to signs and to signals are obviously in need of investigation.

On the general problem of what will or what does the driver do, *i.e.*, "ordinary driver behavior," the following studies may be listed:

1. *Overtaking and Passing Behavior.* A study[5] on the distance and time needed by 80 per cent of the driving population for overtaking and passing on the highway has been reported previously. Interest in this problem was aroused by a report of other investigators that a driver could pass another vehicle in 6 seconds, regardless of speed. However, the conditions of the test indicated that this figure represented what a skilled driver could do under pressure. It was felt that for the practical problem of designing of highways it was necessary to know what the majority of ordinary drivers were able to do in their ordinary performance. Observations were made from a car floating with traffic by means of stop watches and snapshot camera. The resulting distribution showed that the time values ranged from 9 to 12 seconds for speeds from 20 to 50 miles per hour respectively. These figures are median values, and clearly show a relationship to the speed of the vehicle as well as the fact that the general driving public needs more time and distance than the skilled driver.

2. *Judgment Time.* A great deal of work has been done on simple reaction time and choice reaction time in connection with driver tests and traffic safety. In connection with some of the above problems, however, it was felt that both of these, as usually measured, were too simple and that a measure was needed of the time for perception, judgment and reaction on

[5] T. M. Matson and T. W. Forbes, "Overtaking and Passing Requirements As Determined from a Moving Vehicle," *Proc. H'way Res. Bd.*, 1938, 18, 100–112.

the highway. Accordingly, a study[6] was carried out with a series of 20 subjects who at first acted as drivers and later as observers, by a stop watch method. The results are rough, but probably fairly indicative.

Previous studies[7] have shown simple reaction time to be approximately 0.5 seconds, choice reaction time (perception time) to be anywhere from 0.7 to 1.0 second, and the present study of judgment-perception time indicated a median of about 2.2 seconds and an 80 percentile of 2.8 to 3.2 seconds. For practical purposes, it is felt that the 80 percentile value of 2.8 seconds should be used in order to care for the large majority of the drivers on the road. This time value, together with the distance needed for passing, is of importance in determining the minimum length of "tangent" or "straightaway" which should be provided on highways.

3. *Studies of Driver Behavior on Various Designs.* In order to check the effectiveness of present designs and to be able to predict designs which will operate more favorably, it is desirable to have records of the actual behavior of drivers and their cars on the highways under various conditions of speed, curvature, banking, etc. In order to obtain such data, a technique for photographing such driver behavior from directly above certain selected locations has been worked out. (See footnote 6.) This involves the use of a movie camera from a blimp. By means of suitable projection and certain indices giving distance values, the pictures can be measured and the results plotted in terms of distance and speed. Thus it can be demonstrated whether a given curve, ramp, or traffic circle is being used as anticipated by the designing engineers, and furthermore the most convenient and effective path for the driver and the vehicle can be predicted for new conditions of speed, curvature and banking.

[6] T. W. Forbes, "Methods of Measuring Judgment and Perception Time in Passing on the Highway," *Proc. H'way Res. Bd.*, 1939, 19, 218–231.

[7] H. R. DeSilva and T. W. Forbes, *Driver Testing Results.* Harvard Traffic Bureau and W.P.A. of Mass., Boston, 1937, pp. 98.

SUMMARY

1. An application of experimental psychological research methods combined with engineering research techniques is being carried out to obtain data for the guidance of traffic engineers in fitting traffic facilities to the needs and limitations of the driving population.

2. The problems involved fall into two general fields. The first is the question of the limits of ability of the ordinary driver or what the driver *can* do. The second is the question of average or general driver ability, and represents the question of what the driver not only *can* but *will* do.

3. Studies have been carried out on overtaking and passing behavior on the highway, judgment time on the highway, a technique has been worked out for analyzing highway sign efficiency, and studies of legibility of destination sign letters have been completed. Other studies of arrow legibility and of the most convenient or "natural" path of vehicles in the hands of the general driving population are under way.

4. Such psychological or "human engineering" data have been presented in practical form for easy use by traffic engineers in the field. On the basis of such data, the traffic engineer can provide traffic designs and control equipment which will be convenient, efficient and easy for the driver to use. It is felt that the provision of psycho-physiological data of a reliable and practical type for the guidance of the traffic engineer is a positive and promising approach to the problem of both traffic efficiency and traffic safety.

THE RELATIONSHIP BETWEEN SELF-ESTIMATED AND MEASURED VOCATIONAL INTERESTS

RALPH BEDELL
University of Nebraska

INCREASED attention to the significance of the vocational interests of college and university students characterizes a current trend in higher education. Educators have learned that the prediction of student scholastic success on the basis of mental ability and achievement alone fails to consider the psychological satisfactions which a student should obtain from scholastic preparation and later vocational participation. This consideration of the probable happiness or satisfaction which may be experienced in a given vocation has led to the development of various test instruments designed to measure vocational interests.

Those who have used such interest tests for individual counseling of students have noticed that several students show discrepancies between their estimated and measured interests. Such discrepancies always raise the question of the extent to which one should expect estimated and measured vocational interests to differ. It is the purpose of this study to report the relationship between self-estimated and measured vocational interests for a group of freshman women students who definitely had made a vocational choice.

The 141 women students included in this study were all registered as freshmen in the Teachers College of the University of Nebraska. All such freshmen during their first semester take an orientation course, one aspect of which is the provision of clinical counseling to each student who requests such service. As a preliminary to this counseling each student estimated her vocational interests immediately before taking

the Vocational Interest Blank for Women by E. K. Strong.* The comparison between these self-estimated interests and the scores made on this test are the basic data of this study.

The form used for obtaining the estimates is shown in Figure 1. The first paragraph is taken from the directions given

VOCATIONAL INTEREST RATING SCALE

Women

Name ..... Degree for which you are working .....

Major department ..... Number of years you have taught .....

Approximate total number of semester hours earned in college or university ..... Date .....

Indicate after each occupation listed below the extent to which you would like that kind of work. Disregard considerations of salary, social standing, future advancement, etc. Consider only how well you would like to do what is involved in the occupation. You are not asked if you would take up the occupation permanently, but merely the extent to which you would enjoy that kind of work, regardless of any necessary skills, ability, or training which you may or may not possess.

Make a straight vertical mark through the line opposite each occupation at the point which indicates how well you would like that kind of work in comparison with those people who are highly interested and successful in it. It may help if you think of the people you have known who are highly interested and successful in the occupation and compare yourself with how well they like the work.

Work rapidly. Your first impressions are desired here. Rate yourself on each occupation. You may place your mark at any place on the line; it need not be directly under a descriptive phrase.

	How well would you like this occupation in comparison with those now in this work?				
	Would not like	Less than the average	As well as the average	Better than the average	As well as the best
Author					
Librarian					
Artist					
Physician					
Dentist					
Life Insurance Saleswoman					
Social worker					
YWCA Secretary					
Teacher of English					
Teacher of Social Sci.					
Teacher in general					
Lawyer					
Teacher Math-Phys. Science					
Nurse					
Stenographer-Secretary					
General Office Worker					
Housewife					

* Stanford University Press, 1933.

on the Strong Blank, and the occupations are all of those for which this blank may be scored. The method of marking made it possible to refine the estimates to millimeter scores or to group the estimates into five degrees of preference for each occupation.

TABLE 1

Summary of Self-Estimated Interests Made by 141 Teachers College Freshman Women

Vocation	Would not like	Less than the average	As well as the average	Better than the average	As well as the best
Author	56	23	23	21	18
Librarian	25	28	45	33	10
Artist	50	24	26	17	24
Physician	80	22	18	13	8
Dentist	111	19	10	1	0
Life Insurance Saleswoman	110	16	8	7	0
Social Worker	25	17	47	39	13
YWCA Secretary	30	27	40	33	11
Teacher of English	18	25	41	34	23
Teacher of Social Sciences	40	29	44	20	8
Teacher in General	11	14	45	37	34
Lawyer	81	26	22	8	4
Teacher of Math and Physical Sciences	80	21	24	9	7
Nurse	45	37	25	20	14
Stenographer-Secretary	23	20	32	32	34
General Office Worker	30	26	33	36	16
Housewife	8	3	38	32	60

Table 1 shows the number of the 141 women who estimated their interests in each of the five categories for each occupation. Examination of this table shows a wide variety of vocational interests. Although interests in teaching tend to predominate, there is no evidence that this group estimates themselves to be homogeneous with respect to a "like for" teaching or any of the other occupations.

The ratings derived from the scores on the *Vocational Inter-*

est Blank shown in Table 2 reveal a lack of homogeneity similar to that found for the estimates. A ratings, which indicate a high degree of measured interest in the vocation,

TABLE 2

Summary of Ratings on the Strong Vocational Interest Blank for Women Made by 141 Teachers College Freshman Women

Vocation	C	B	A
Author	77	55	9
Librarian	105	36	0
Artist	76	54	11
Physician	130	11	0
Dentist	134	6	1
Life Insurance Saleswoman	120	20	1
Social Worker	129	12	0
YWCA Secretary	129	11	1
Teacher of English	87	45	9
Teacher of Social Sciences	117	17	7
Teacher in General	74	53	14
Lawyer	124	16	1
Teacher of Math and Physical Sciences	94	42	5
Nurse	23	91	27
Stenographer-Secretary	4	59	78
General Office Worker	2	72	67
Housewife	2	38	101

are much less numerous than the C ratings. While housewife is the only occupation which shows a majority of A ratings, there are a fair number of such ratings for stenographer-secretary and general office worker. Still fewer A ratings are found for nurse, teacher in general and artist. A ratings for any other single occupation are found for less than six per cent of the subjects. Ratings, in general, appear to be lower than the estimates. This probably means that the subjects estimate themselves to be more mature than the Blank measures them to be.

The relationships between self-estimates and scores are shown in Table 3. The Pearson coefficients of correlation were computed between the millimeter measures of estimates

TABLE 3

Coefficients of Correlation between Self-Estimated Interests and Scores on the Strong Vocational Interest Blank for

TABLE 3

Coefficients of Correlation between Self-Estimated Interests and Scores on the Strong Vocational Interest Blank for Women for 141 Teachers College Freshman Women

Scores	Self-Estimates															
Vocation	Author	Librarian	Artist	Physician	Dentist	Life Insurance Saleswoman	Social Worker	YWCA Secretary	Teacher of English	Teacher of Social Sciences	Teacher in General	Lawyer	Teacher of Math.-Phys. Sci.	Nurse	Stenographer-Secretary	Office Worker	Housewife
Author	.29	−.15	.09	−.13	−.17	−.15	−.19	−.31	−.07	−.09	−.13	−.05	−.25	−.01	−.34	−.31	−.09
Librarian	.11	−.04	−.01	−.19	−.17	−.15	−.24	−.22	.07	.07	−.06	−.12	−.07	−.06	−.27	−.15	−.05
Artist	.06	−.20	.36	−.15	−.16	−.16	−.25	−.38	−.15	−.25	−.16	−.14	−.12	.08	−.31	−.25	−.20
Physician	−.16	−.21	−.17	.34	.15	−.21	−.23	−.26	−.17	.01	−.35	−.01	.27	.12	−.51	−.33	−.25
Dentist	−.33	−.10	−.08	.35	.19	−.20	−.20	−.17	−.21	−.09	−.01	−.04	.37	.16	−.25	−.12	−.13
Life Insurance Saleswoman	.12	.04	.14	−.15	−.19	.08	.12	.17	.01	−.04	.02	.05	−.16	−.13	.01	−.01	.09
Social Worker	.09	−.07	.05	.15	.10	−.09	.28	−.09	.07	.17	.09	.09	.06	.10	−.36	−.36	.21
YWCA Secretary	.40	.16	−.17	−.02	−.02	.02	.34	.28	.29	.30	.19	.00	.12	.00	−.09	−.09	.20
Teacher of English	.54	.11	.02	−.10	−.08	−.09	.15	.03	.39	.23	.26	−.01	−.12	−.04	.03	−.07	.28
Teacher of Social Sciences	.28	.13	−.01	.03	.08	−.03	.25	.10	.51	.51	.29	.08	.14	−.01	.08	−.04	.12
Teacher in General	.32	.00	.02	.09	.00	−.03	.09	−.01	.42	.36	.36	−.01	.19	−.01	−.03	−.06	.04
Lawyer	.30	.00	.01	.04	−.11	−.03	.08	−.04	.12	.14	−.05	.29	.10	−.06	−.17	−.23	−.03
Teacher of Math.-Phys. Sci.	−.24	−.02	−.15	.34	.11	.04	−.07	.04	.08	.14	.12	−.06	.44	.08	.02	.09	−.08
Nurse	−.03	.04	.13	.12	.05	−.32	.12	.01	−.08	−.14	.10	−.10	.15	.03	.02	−.01	.04
Stenographer-Secretary	−.11	.14	−.04	−.02	.01	.13	.07	.20	−.16	−.22	−.12	.07	−.07	−.04	.54	.46	.19
Office Worker	−.24	.15	−.14	−.06	−.04	.16	.05	.26	−.08	−.24	−.14	.09	−.03	−.08	.53	.48	.03
Housewife	−.21	.25	−.39	−.22	.30	.14	.23	.23	.02	−.17	.10	−.13	−.03	.04	.44	.42	.23

and the raw scores obtained from the Strong Blank for all occupations. The coefficients in the upper left-lower right diagonal show the extent to which self-estimated interests in any given occupation may be used to predict scores obtained from the Blank for that occupation. While thirteen of the coefficients exceed .22, which is the one per cent significance level coefficient, none of these thirteen is high. Only the coefficients for stenographer-secretary and teacher of social sciences exceed .50. Obviously, prediction based on such low coefficients is not sufficiently accurate to assume that self-estimated interests in these occupations can be used to determine interests as measured by the Blank.

Examination of the columns in Table 3 reveals that many self-estimates are no more closely related to scores for the occupation estimated than to the scores for some occupation other than the one estimated. Considering that negative correlations are as valuable for prediction as positive ones, examination of the columns shows that in general self-estimates predict scores for the given occupation no more accurately than such estimates predict the scores for some other occupation. The only exceptions are teacher of social sciences, lawyer, teacher of mathematics and physical sciences, and office worker. Examination of the rows reveals that the scores for many of the occupations are no better bases for predicting self-estimates for the given occupation than for some other occupation. The only exceptions are teacher of mathematics and physical sciences and stenographer-secretary.

It is clear from the data in Table 3 that within the limits of the occupations measured by the Strong Blank there is no close relationship between any self-estimated vocational interest and any measured vocational interest shown by this group of Teachers College freshman women. It is difficult to conceive any pattern of meaningful relationships which could be used as a basis for predicting scores from estimates.

The question of the relative validity of self-estimated and measured interests is a logical result of the study. Although

no evidence reported here can be brought to bear on the problem, it seems to the writer that measured interests are, in general, far more nearly valid than the self-estimates. It is doubtful if college freshmen have had sufficient experiences with occupations of this type to permit valid estimates. The keys with which the Strong Blank is scored are based upon a comparison between the markings of mature women successful in the given occupation and those of a group of mature women representative of the general female population. The responses of such mature women successful in the given occupation probably approximate more closely a valid standard of comparison than college freshmen estimates formed from limited occupational experience.

Some may argue that the possession of likes and dislikes similar to those of women successful in a given occupation (measured) is not necessarily the same as the possession of interest in the occupation (estimated), consequently high correlation between the two should not be expected. The writer now has in progress a study on mature women that may yield evidence on this point. It seems, however, that the more probable explanation lies in the maturity of vocational experiences of the subjects. Such immaturity would tend to invalidate the estimates and to reduce the scores made on the Blanks. It may also tend to invalidate the results obtained from the Blank, but probably not to the same extent as with the estimates. Any of the above would tend to reduce correlations in the general direction of those found in this study.

Consideration of the conclusions from this study in the light of these facts lends increased support to personnel workers who insist that students' self-estimated vocational interests are insufficient evidence upon which to diagnose the amount of satisfaction that will be obtained in a vocation. Such workers must consider both estimated and measured interests in the light of complete case data if an adequate diagnosis is to be reached. A grave question is raised toward the validity of educational procedures largely based upon the self-estimated interests of students.

SUMMARY

The problem of this study was to determine the relationship between self-estimated and measured vocational interests. The self-estimated and measured interests for the occupations included in the keys for the Strong's *Vocational Interest Blank* as obtained from 141 University of Nebraska Teachers College freshmen women constitute the basic data for this study. The group is shown to be heterogeneous with respect to any of the self-estimated or measured vocational interests. Pearson coefficients of correlation are presented between the self-estimates for each occupation and every raw score obtained from the Strong Blank. No high correlations were found. Self-estimates for all occupations except four were found to offer no better bases for the prediction of scores in the given occupation than for the prediction of scores in some other occupation. Scores for all occupations except two were found to offer no better bases for the prediction of self-estimates for the given occupation than for some other occupation. No close relationships between self-estimated and measured vocational interests were found. The writer argues that measured interests are more nearly valid than self-estimates, although the latter are not necessarily without significance. Consequently personnel workers should compare students' measured vocational interests with self-estimated interests. Educational procedures largely based on the self-estimated interests of students are questioned.

THE EFFECT OF LINE-LENGTH ON READABILITY

MATTHEW LUCKIESH AND FRANK K. MOSS
Lighting Research Laboratory, General Electric Company, Cleveland

THIS study of the influence of line-length upon readability is based upon data obtained by the criteria of (1) rate of involuntary blinking, (2) the character of the eye-movements in reading and (3) the rate of normal reading. The techniques of applying two of these criteria have been previously described in detail.[1] An analysis of the eye-movements during reading is introduced as a criterion in our typographical studies for the first time. The application of this criterion is now practicable due to the development of suitable apparatus for amplifying and recording the currents produced by the movements of the eyes without the introduction of instrumental distortions. In approaching this problem from these experimental viewpoints, it is necessary to assume that the subjects, whose reactions and responses are studied, are typical of readers in general. For present purposes, we selected at random, from technical and clerical groups, twenty-four adult subjects possessing so-called normal vision who read under as near normal conditions as may be obtained in the laboratory.

The text-matter read during the present investigation consisted of chapters from the "Outline of History" by Wells. This material was set in 10-point Textype with 2 points of leading and composed in line-lengths of 13, 17, 21, 25 and 29 picas, respectively. The printing was in black ink on non-glossy white paper. In all cases, the reading was done under a constant and diffused illumination of 10 footcandles. Further-

[1] Matthew Luckiesh and Frank K. Moss, "Frequency of Blinking as a Clinical Criterion of Ease of Seeing," *Am. J. Ophthalmol.*, 22, 1939, 616. "Effects of Leading on Readability," *J. Appl. Psychol.*, 22, 1938, 140.

more, the subjects were instructed to read at their normal or customary rates and in addition, every effort was made to minimize unnatural reactions on the part of the subjects. The fact that our subjects have been used repeatedly in visual researches facilitated the attainment of this desirable, if not essential, attitude and reaction toward the research procedures.

Rate of Blinking. The data of Table 1 and Figure 1 sum-

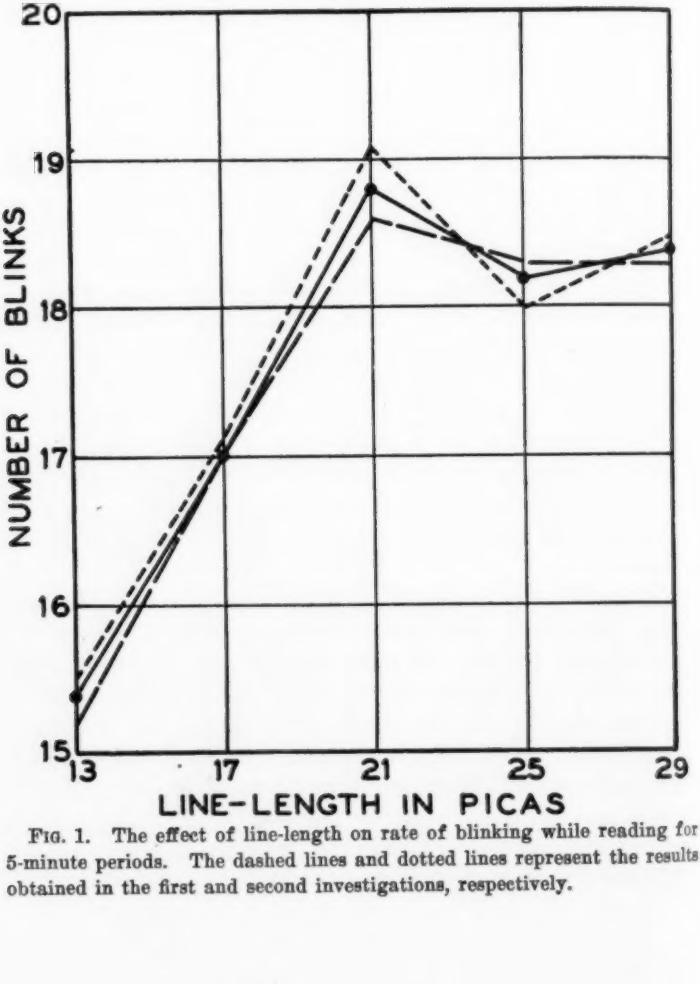

FIG. 1. The effect of line-length on rate of blinking while reading for 5-minute periods. The dashed lines and dotted lines represent the results obtained in the first and second investigations, respectively.

marize the average rates of blinking of the 24 subjects, each of whom read each of the five materials for 5-minute periods on two different occasions. An adequate supply of each of the five materials was available for the total reading-time of 10 minutes in each of the two series of investigations.

TABLE 1

The Geometric Average Rates of Involuntary Blinking During 5-Minute Periods of Reading. The Corresponding Probable Errors Are Relative Values Which Exclude Systematic Differences in Blink-Rate among Individual Readers. (A Pica Equals One-Sixth of an Inch.)

	Line-length in Picas				
	13	17	21	25	29
Number of blinks					
First series	15.2	17.0	18.6	18.3	18.3
Second series	15.5	17.1	19.1	18.0	18.5
Both series	15.4	17.0	18.8	18.2	18.4
Relative blink-rate	100.0	110.4	122.0	118.2	119.5
Relative probable error		3.4	3.8	3.3	3.3

The outstanding fact revealed by these data is that the fewest number of blinks occurred during the reading of the shortest lines. Since a reduction in the rate of involuntary blinking has been shown to result from favorable changes in the visual situation,[2] including such factors as illumination, type-size, refractive errors, glare, duration of the task and others, the data of Table 1 indicate that the shorter lines are read more easily than the longer lines. Such a conclusion is inescapable unless it be assumed that rate of blinking is not an appropriate criterion of readability in this particular visual situation. In the light of present knowledge, however, there does not appear to be any logical or established basis for such an assumption.

These data also indicate (1) that the readability of these materials decreases as the line-length is increased from 13 to

[2] Matthew Luckiesh and Frank K. Moss, "A Summary of Researches Involving Blink-Rate as a Criterion of Ease of Seeing," *Illuminating Engineering*, 35, 1940, 19.

about 21 picas and (2) that no important changes in readability result from increases in line-length from 21 to 29 picas. Thus it appears that the line-length used in most books is not the most favorable one from the viewpoint of readability. While it is recognized that many so-called practical factors might oppose the use of lines as short as 13 picas, the indication that such a line-length is superior to longer lines may be an important factor in guiding the trend of typographical practice. This interpretation of the present data is similar to that advanced with respect to the factor of boldness of type-face.[3] For example, our studies indicated that type-faces which are bolder than those commonly used are desirable from the viewpoint of readability. This fact may be of considerable importance in guiding typographical practice even though the optimum degree of boldness indicated by these researches might not be recommended immediately.

The indicated superiority of 13-pica lines, as compared with longer lines, is particularly interesting from the viewpoint of newspaper typography since this line-length is approximately the same as that in common use. Although it seems reasonable to extend the interpretation of the present data to newspaper typography, it would be advisable to confirm these conclusions by researches involving actual newspaper materials, including multiple columns on the same page.

In view of the influence of line-length on readability indicated by the data of Table 1, it is important to appraise the statistical reliability of the differences in rate of blinking during the reading of these materials. In general, it will be noted from Table 1 that the average rates of blinking for the first series of 5-minute reading periods are substantially the same as those for the second series in which the order of reading the five materials was reversed. This fact is considered particularly significant since a week or longer intervened between the two series of reading-periods. More specifically, it will be noted that the rate of blinking while reading the 21-pica lines

[3] Matthew Luckiesh and Frank K. Moss, "Boldness as a Factor in Type-Design and Typography," *J. Appl. Psychol.*, 24, 1940, 170.

is 22 per cent greater than that for the 13-pica lines. Since this difference is five times as large as the probable error of the difference, the probability that it is an accidental one is about one chance in 1500. From a similar viewpoint, the indicated superiority of 17-pica lines, as compared with 21-pica lines, is also considered to be statistically reliable. However, the differences in blink-rate among the 21, 25 and 29 pica lines are not particularly significant, not only because they are small in absolute value but also because they are small in comparison with their corresponding probable errors.

The data of Table 2 summarize the behavior of the individual subject while reading the various materials. In view of the human vagaries involved in reading, and in view of the many causes of blinking, the agreement among the individual subjects with respect to the superiority of the shorter lines is considered excellent. These data also confirm the conclusion that the 21, 25 and 29-pica lines are about equally readable.

TABLE 2

Showing the Number of Subjects Who Blinked Faster, the Same or Slower While Reading 29-Pica Lines Than in Reading 13, 17, 21 and 25-Pica Lines, Respectively

Line-length	Number of Subjects		
	Faster	Same	Slower
13 picas	18	2	4
17	16	.....	8
21	11	1	12
25	13	1	10

In interpreting the data of Table 1 from theoretical viewillusion that the shorter lines were leaded more than the longer ones. This fact, in itself, appears to be significant. It is not improbable that this illusion arises from differences in the points, the illusion of variable leading and the actual variable of the number of hyphenated words are factors which warrant consideration. Nearly all of the subjects experienced the angular deviation, due to line-length, as the point of visual

fixation sweeps from the end of one line diagonally downward to the beginning of the next line. For example, it is a well-known fact that ocular muscular movements play an important part in the introspective estimation of distance. Thus as the length of the lines is decreased, the angular deviation of the return sweeps, increases, and the separation between the lines *appears* to increase. It is conceivable that the apparent increase in leading as the line-length is reduced may be "equivalent" to an actual increase in leading when considered from the physiological basis of muscular functions and from the psychological viewpoint of the distraction of adjacent lines of print. In any event, the illusion of greater leading should be a favorable rather than an unfavorable factor in readability.

It is possible that the greater number of lines ending in hyphenated words in a given amount of material composed in the shorter lines might be a factor related to readability and one which might influence the rate of blinking while reading. An analysis of all of the material used in the present investigation yielded the following data:

Length of line	13	17	21	25	29
Hyphens per pica	0.020	0.015	0.006	0.005	0.005
Hyphens per line	0.25	0.26	0.12	0.13	0.14

If the frequency of hyphenated words is a significant factor, it seems reasonable to assume that the rate of blinking while reading the shortest line would be augmented to the greatest extent. Thus the conclusions following from the data of Table 1 appear conservative.

Eye-movements. Notwithstanding the fact that the brain demands accurate and coordinated muscular responses in critical seeing (thereby minimizing the influence of unfavorable visual factors) it seems reasonable to assume that the ocular movements would not be executed equally well in diverse situations. However, it was also recognized that possible differences in the movements of the eyes would not be marked unless the visual mechanisms were fatigued from long periods of exacting work. Thus the application of the criterion of

eye-movements in reading involves the recording of these movements in detail and under conditions which, in themselves, do not place the subject under abnormal restraint. In our procedures, the eye-movements were recorded by means of the electrical currents arising from such movements. This was accomplished by placing small electrodes on the face and in the region of the right eye from which the feeble currents created by the eye-movements were led from the subject, through an amplifier and recorded by means of an oscillograph. This apparatus has been so developed that extensive statistical studies of the eye-movements are practical.

In analyzing these records it was possible to determine the number of fixations per line with considerable certainty. In addition, certain unique characteristics of the tracings of the eye-movements during the reading of lines of different lengths were observed. However, no attempt has been made to present an analysis of these characteristics since the movements are irregular and complex. The data of Table 3 summarize the number of fixations per line for line-lengths of 13, 21 and 29 picas, respectively. These data were obtained from sections of the records which indicated steady or rhythmic reading with clearly defined fixational-pauses.

The data of Table 3 show that the average number of fixations increases as the line-length increases, but not in direct proportion. Thus relatively fewer fixations per pica of line-length occurred during the reading of 29-pica lines than for

TABLE 3

These Data Were Determined from the Records of 12 Subjects, Each of Which Included the Reading of 15 or More Lines of Print. The Probable Errors of the Data Relating to Length of Fixation-Span, Including Systematic Differences among Subjects, Are Less Than 2 Per Cent.

	Line-length in picas		
	13	21	29
Number of fixations	3.92	5.84	7.64
Fixation-span in picas	3.32	3.60	3.80
Relative length of span	100.0	108.5	114.5

13-pica lines. Conversely, the average fixation-span increased as the line-length increased. For example, the average fixation-span increased from 3.3 to 3.8 picas or from 8 to 9 ten-point characters when the lines were increased from 13 to 29 picas. Since the increase in fixation-span could scarcely be attributed to cortical control, it seems reasonable to assume that it is initiated by physiological phenomena. Certain theoretical considerations support this supposition. For example, let it be assumed that the vision is directed to the center of a given line of print when the eyes are in a so-called normal position. Obviously as reading progresses, the deviation of the point of fixation from this assumed median position is greatest when reading the longest lines. Hence it might be expected, from a strictly mechanical viewpoint, that the muscular mechanisms would become less capable of making small and precisely controlled steps in fixation as the line-length is increased. This theoretical assumption is supported by the data of Table 3. It may also be assumed that reading would become somewhat more difficult as the eye-movements became more erratic.

Although the criteria of blink-rate and character of the eye-movements indicate that the shorter lines are more readable than the longer ones, the correlation between the two sets of data is not perfect. For example, the data of Table 1 indicate but little change in readability for line-lengths greater than 21 picas whereas the data of Table 3 indicate substantial differences. However, it is obvious that the frequency of fixations is but one characteristic of the eye-movements and hence it should not be interpreted as an adequate criterion of readability. On the other hand, the rate of involuntary blinking may be considered as such a measure since it may be influenced, directly or indirectly, by all factors involved in the act of reading.

Speed of Reading. The geometric average rates of reading the various line-lengths, as determined from the individual rates of 24 subjects who read for 5-minute periods on two occasions, are shown in Table 4.

TABLE 4

	Line-length in picas				
Picas read per minute	13	17	21	25	29
First series	601	596	580	580	594
Second series	590	562	573	588	593
Both series	596	579	576	584	594
Relative rate	100.0	97.2	96.7	98.0	99.6

It will be noted that the maximum difference in rate of reading is about 3 per cent. Thus our contention that rate of normal reading is, at best, an insensitive criterion of readability is again confirmed.[4] These results indicate (1) that the shortest and longest lines, respectively, were read at about the same rate and (2) that the slowest rate of reading occurred with the 21-pica lines. Thus it is not probable that the influence of line-length upon speed of reading can be related to some one variable such as frequency of the return sweeps, apparent degree of leading, etc. In general, these data suggest a somewhat vague qualitative agreement with the results obtained with the criterion of rate of blinking with respect to the superiority of the shorter lines.

Conclusions. These studies indicate that the readability of 10-point Textype with 2 points of leading diminishes as the length of line is increased from 13 to about 21 picas; and that further increases in line-length result in less important changes in readability. Although lines of 13 picas in length are revealed as superior to lines of greater length, it should not be assumed, for example, that two 13-pica columns on the same page would be superior to a single column 26 picas in measure in view of the possibility that the second column might be a distraction in the reading of the first, and *vice versa.* Further research may confirm the former assumption, but it should not be made in the absence of specific data obtained under the specified conditions.

[4] Matthew Luckiesh and Frank K. Moss, "Criteria of Readability," *J. Exp. Psychol.*, 27, 1940, 256.

THE SELECTION OF DEPARTMENT STORE PACKERS AND WRAPPERS WITH THE AID OF CERTAIN PSYCHOLOGICAL TESTS

MILTON BLUM
College of the City of New York
AND
BEATRICE CANDEE
Vocational Service for Juniors

A LARGE department store hiring many packers and wrappers for the Christmas season asked the New York State Employment Service to investigate the use of tests for this purpose. They agreed to have the employees who had already been engaged on such jobs tested so that the validity of the test results could be checked with this group as well as with inexeperienced workers.

Fifty-three permanent employees were given the O'Connor Finger Dexterity test, the Minnesota Placing Test and the Minnesota Turning Test. All tests were administered by an experienced examiner. The employees were tested in a specially designated room at the store. They were informed by both management and the psychologist that the test results were in no way to be considered as a check on their ability, but were being given to see if such tests could select future employees. Apparently there were no unfavorable emotional attitudes on the part of the subjects.

The second group of subjects consisted of one hundred thirty people who were given the three tests at the New York State Employment Service. Of this number, thirteen were eliminated on the basis of exceptionally poor test results since the store felt that this should be a service as well as a research project. The employment service ultimately made ninety-two

referrals to the store. All but three, who for some reason did not meet the requirements of the store, were hired. This group will hereafter be referred to as the seasonal employment group.

The permanent employees tested worked as packers and wrappers, relief cashiers, cashiers, and assorters. Production records were available for the thirty-six employees engaged on the first two mentioned jobs only. Supervisor's ratings were available for thirty-eight of these employees. The average production for a six months' period for the relief cashiers was 12,691 units. The packers and wrappers achieved an average production record of 25,759 units. This meant that the different jobs offered different opportunities. Of the eighty-nine seasonal employees hired, only 52 subjects were employed on jobs which could be measured on a production basis. Supervisor's ratings were available for forty-four of these subjects.

Table 1 presents the average test score for each group on the various tests. The test results for the various employee groups show that the new workers when selected had slightly inferior test scores than the permanent employees. Only on the Placing test, however, is the difference between the averages statistically significant. In other words, with the elimination of only one-tenth of their referrals on the basis of tests, the interviewers of the employment service had selected a group only slightly inferior in test performance to the permanent employees on these jobs. This in itself is interesting. Four years before this when using the Finger Dexterity to select similar workers in another store, we found that in five weeks' time, interviewers developed a 97 per cent accuracy in selecting applicants to pass a critical score on the test.

The intercorrelations of the tests based upon a sampling of one hundred thirty females seeking the positions open were as follows:

r. between Finger Dexterity and Placing test was $+.416 \pm .07$
r. between Finger Dexterity and Turning test was $+.335 \pm .07$
r. between Placing and Turning test was $+.551 \pm .06$

The attaining of an acceptable criterion of success was, as

TABLE 1

Average Score on Tests for Seasonal and Permanent Employee Groups

	No.	F.D. aver.	F.D. σ	Plac. aver.	Plac. σ	Turn. aver.	Turn. σ
Seasonals hired as wrappers	27	7′ 48	46″	220″	16″	175″	19″
Seasonals hired as packers	10	7′ 41	43″	225″	11″	174″	13″
Seasonals hired as wrap-cashiers	15	7′ 56	57″	225″	9″	178″	13″
Permanently employed relief cashiers	11	7′ 24	52″–	205″	15″	166″	17″
Permanently employed wrap-cashiers	25	7′ 20	47″	201″	12″	167″	23″

usual, a most difficult task. The most tenable criterion was the actual production record of each employee. During the month of December all employees are probably working closer to maximum ability than at any other time of the year. It was, therefore, decided to consider the average daily number of packages wrapped during this month, as the criterion.

This criterion is acceptably reliable. The reliability coefficient using the split-half technique, when average production for the first half of the month of December was compared with the average production of the second half, was + .88. Even when different seasons of the year are compared, a correlation (rank-difference method) of + .70 was obtained between the average daily production record for the group of thirty-six permanent employees over a six month period and the average daily production record for those employees during the month of December.

The average daily production in December for each of the groups investigated is presented in Table 2.

The permanent employees seem to be clearly superior on the basis of production records. On the jobs which are directly comparable, those of wrapper-cashier, their performance is more than twice as good. This difference between the groups

is much greater than shows on the test scores and might indicate either that experience largely determines the better performance of the permanent workers or that these particular tests do not get at the abilities involved. An additional factor

TABLE 2

Average Daily Production During December

	Average daily production	σ
Seasonals employed as wrappers	102	27
Seasonals employed as packers	111	20
Seasonals employed as wrap-cashiers	79	17
Permanently employed wrap-cashiers	172	76
Permanently employed relief cashiers	91	43

probably is that the permanent workers are given preference for the more productive jobs.

Table 3 presents correlations existing between the criterion, production records, and the various test results.

TABLE 3

Correlation between Production Record and Test Results

	Plac.	Turn.	F.D.
Seasonal employees production record	.35	.27	.02
Permanent employees production record	.21	.06	.08

While the 10 per cent of the distribution which was eliminated would probably reduce the correlation coefficient, it is not likely that the relations would have become meaningful if it were included in the case of either the Turning or the Finger Dexterity.

It is interesting that both the Placing and Turning give higher correlations with the new workers than with the permanent group, which apparently indicates that even the slight predictive value possessed by the tests is reduced by experience on the job.

The relation existing between the Finger Dexterity test and production record for both groups is zero. This result is rather surprising in view of the fact that the Finger Dexterity test has had much prestige among department stores in New York City in the selection of packers.

The Placing test yields better results. However, although the correlations with the Placing test are as high as those frequently reported, they would in no way justify the selection of workers on the basis of this test alone.

The multiple correlation between production records of the seasonal employees and all three test scores was + .38 and + .24 between production records of the permanent employees and test results. This indicates that the 3 tests together give only slightly better results than the Placing alone. A relationship is present for the seasonal workers, which is reduced by experience as in the case of the single tests. However, even with the tests combined the correlations between production records and test results were too low to be valuable for prediction in individual cases. It was decided to compare the various groups when divided according to high production and low production records. (High defined as the highest third; low defined as the lowest third). The average score for each of the groups mentioned in Table II was then computed. Naturally, when subdivided to this extent, the numbers of cases in each comparison is small and does not warrant the actual presentation of the average time to complete the test and the standard deviation. However, the comparisons were in the expected direction (average time of highest third in production record faster than average time of lowest third in production record) in 11 of the 15 instances. Like the correlations, this method would indicate that the tests are not valueless but that the test results are not good enough when individual selection is the problem to be solved. However, when mass hiring for seasonal work is necessary, such testing may be of some aid to the personnel department provided it is worth the cost of administering the tests, and providing a serviceable critical score can be found for the specific store.

One additional criterion, supervisor ratings, was available. The ratings obtained in this study were made by supervisors as a routine procedure. They rate all employees with the following consideration in mind:

Rate A:—all who are to be continued in employment.
Rate B:—all who are good and should be re-hired at the first opportunity available.
Rate C:—all who are to be re-hired for seasonal work.
Rate D:—all who are inefficient.

TABLE 4

Distribution of Supervisor's Ratings

	N	A	B	C	D
Seasonal employees	44	2%	52%	46%	0%
Permanent employees	38	82%	13%	5%	0%

Here the permanent employees are apparently clearly superior. However, these ratings require further explanation because of the policy of the Personnel Department of the store. Since most permanent employees are to be continued in employment, and since with very rare exceptions all seasonal workers are laid off after the holidays and rehired later as the opportunity occurs, a "B" rating for a seasonal worker may easily be equivalent to an "A" for a permanent employee. However, the comparison cannot even be made simply on this basis because the practice of defining an "A" as designating a person to be continued in employment undoubtedly eliminates practically all differences among the permanent employees except the distinction between those who are actually unsatisfactory and those who are acceptable.

A comparison of the test scores of the various groups in Table V shows that the permanent employees with "A" ratings obtain average scores on all three tests that are superior to the average test scores of all seasonal employees, both those

TABLE 5

Comparison of Test Results for Various Groups

Finger dexterity	N	Average	Range	σ	σm
Dept. store perm. employees-rating "A" (Gr. 1)	31	436	366–565	46.8	8.32
Permanent employees below "A" (Gr. 2)	7	457	375–584	57.1	21.92
Seasonal employees "Above C" (Gr. 3)	24	475	407–553	41.5	8.47
Seasonal employees "Rated C" (Gr. 4)	20	476	362–624	54.7	12.23
Queens N.Y.S.E.S. group—(Gr. 5)	420	463	315–735	69	3.38
Placing					
Gr. 1		201	174–230	13.2	2.40
Gr. 2		209	189–231	13.2	5.07
Gr. 3		232	192–260	15.5	3.16
Gr. 4		226	198–247	10.8	2.44
Minn. sample*		228	140–340		
Turning					
Gr. 1		167	141–269	23.3	4.23
Gr. 2		163	138–189	16.7	6.29
Gr. 3		183	135–221	16.1	3.29
Gr. 4		185	154–205	15.4	3.44

* Dvorak, B. J. "Differential Occupational Ability Patterns"—University of Minn. Research Instit., Vol. 38, 1935.

rated "C" and those rated "above C." The $D/\sigma_{diff.}$ in all cases is greater than three with one exception. (The $D/\sigma_{diff.}$ between average F.D. score for rating of "A" group and Rating of "C" group = 2.7.) The $D/\sigma_{diff.}$ between the permanent employed group obtaining an "A" rating and both seasonal groups on the placing test is greater than 7.

When the permanent employees rated below "A" are compared with the seasonal group rated above "C" the $D/\sigma_{diff.}$ is above 3 on the Placing, but 2.8 on the Turning and .7 on the Finger Dexterity. These results indicate that the "less satisfactory" permanent employees are better than the "above

average'' temporary employees on the Placing test. They tend to be better on the Turning test and possibly also on the Finger Dexterity test. This is still further indication that the store experience seems to raise the test score to some extent, even though the tests do not predict better workers within either the more or less experienced groups. In all comparisons between the two seasonal groups and between the two permanent groups, no statistically reliable differences are found on any tests.

When the employees of the store are compared with available control groups, the permanent employees with "A" ratings clearly excel on the Finger Dexterity test a group of 420 women applying for dexterity jobs at the State Employment Service, both in average score ($D/\sigma_{diff.} = 3$) and in the elimination of the lower range. The good seasonal group, however, is no better in average score on either the Placing or the Finger Dexterity than the controls, although about 8 per cent in the lowest range are eliminated. There is no way of knowing whether these people in the lowest range of the test scores would also have been satisfactory workers or whether among them would have been found employees so unsatisfactory that they could not be retained through the season. Another department store in New York in a study some years ago on the Finger Dexterity found that by using a much higher critical score than this it could have eliminated 10 out of 14 or 71 per cent of the new workers discharged before the season was over. In so doing it would also have eliminated 29 out of 47 or 61 per cent of its acceptable workers and one out of three or 33 per cent of its best seasonal employees. It seems most improbable that we did eliminate all of the poor workers in the lowest 8 per cent since the other study, which had a better test prediction in all ranges, still would have eliminated a very small proportion of its bad workers on the basis which we used. It therefore appears that a genuine difference exists in the results of the two studies, due to a difference in the criteria used, in the handling of the workers, in the store policy, or some

other factors in the general situation. It must also be noted that even the more favorable results of the other study show so high an elimination of acceptable and even excellent workers along with the less desirable ones that the use of such a method of selection by a public employment service would be questionable even though an individual store may adopt a method of this kind if it has a very large labor supply.

In our study the complete absence of any workers rated as inefficient even in the seasonal group is to be noted. The store had considered the selection of workers so good that it cut a full day from the training period. If the selection actually was much better than previous seasons, it must be attributed to other factors than the tests. It is recognized that the interest and attitudes fostered by such a study may in themselves bring about more favorable production records.[1]

To summarize: In this study the satisfactory permanent employees excel less experienced workers in production, in supervisor's ratings and in test scores, the indications being that this is due primarily to experience on the job. That experience rather than selection is the chief factor is indicated by the fact that there are no significant differences on the tests between the satisfactory and less satisfactory permanently employed groups and between the two seasonal employed groups but only between the more and less experienced workers, regardless of the supervisor's ratings. Also the correlations between the production records and the scores on the tests are noticeably lower in the experienced group. Thirdly, while the satisfactory experienced workers excel the general population on the Finger Dexterity and the Placing, the seasonal workers do not excel in either test although the store considered them an exceptionally good group of new employees.

On the basis of these results the only justification for the use of these three tests in the selection of department store wrap-

[1] This was clearly shown in the Hawthorne Study, Western Electric Company. Elton May, "Human Problems of an Industrial Civilization," Macmillan Co., 1933, pp. 194.

pers and packers, is if the hiring of seasonal workers is viewed as a mass selection with little attention paid to individuals or to other types of information about the applicant. If much consideration is to be given to individuals as prospective permanent employees, then individual test scores for any one person should not be considered too seriously. The job of packing and wrapping is a relatively simple one and can be learned within a short period of time as is indicated by the fact that the correlation existing between production of inexperienced people and test scores are reduced with experience.

This study seems to indicate, as do most others on dexterity tests, that the specific factors in dexterity functions considerably outweigh any more general ones. Apparently experience in wrapping does have a slight effect in raising test scores on three different tests and in reducing initial differences among the workers on the tests. However, on the job, specific factors seem to outweigh any which exist in common with either of the three tests since none of them will select good workers from the population as a whole, unless they are already experienced. Another study now in progress is being conducted in a different department store to check these results.

MALADJUSTMENT IN COLLEGE, PREDICTED BY BERNREUTER INVENTORY SCORES AND FAMILY POSITION

WILLIS FISHER AND SAMUEL P. HAYES, JR.[1]
Sarah Lawrence College

FOR a number of years, the Bernreuter Personality Inventory has been one of a battery of tests administered to entering students at Sarah Lawrence College. The individualized methods of instruction and personal guidance utilized there make it more likely than at most colleges that any existing cases of serious maladjustment will be discovered by the faculty and referred to the college psychiatrist. Consequently, an unusual opportunity is afforded for evaluation of the usefulness of the Inventory to college authorities in drawing early attention to possible cases of maladjustment, either for the purpose of admitting "good risks," or for the purpose of indicating special precautions to be taken during the college course.

In order to secure an independent criterion of maladjustment which could be used to validate the Bernreuter Inventory, the college psychiatrist was asked to indicate which students, on a list of 438, he considered to have had personality difficulties serious enough to prevent a successful adaptation to the college, or to render doubtful the value of college experience to them in making an adaptation later, or both. All of these 438 students had been in the college for some period during the psychiatrist's four-and-a-half years

[1] Mr. Fisher obtained the data on the Inventory scores and the criterion. Mr. Hayes obtained the material on family position, did the statistical analysis, and prepared the article in its present form. Both contributed to the interpretation and conclusions. They wish to express their obligation to Marjorie Frost and Judith Schoellkopf for statistical help.

of continuous service there, although some of them had been there for periods as short as one or two semesters.

This criterion was not perfect, of course. Its reliability was impaired by differences in the ability of teachers to recognize serious personality disturbances, in the social impact of such disturbances, in the length of time spent at the college, and differences of opinion about the desirability of sending specific cases to a "psychiatrist," with the stigma sometimes attached to that idea. On the other hand, while some serious cases of maladjustment were undoubtedly overlooked, it was obvious that the students indicated by the psychiatrist all had serious personality problems.

Twenty-five students were in the group indicated by the psychiatrist as seriously maladjusted.[2] When their scores on the Bernreuter Personality Inventory were compared with those of the other 413 students, the mean score of the 25 was higher than the mean score of the others on all of the scales except B4–D. These five differences between means were all statistically reliable. On the B4–D scale, the mean of the 25 was lower than the mean of the others, but the difference was not great enough to be considered statistically reliable. The relationship between scores on these scales and "serious maladjustment" can be expressed by a coefficient of biserial correlation (which is based on differences in the means of the two groups), as in the following table.

Since the coefficients for scales B1–N, B3–I and F2–S were high enough to be of possible predictive significance, the question arose of selecting a score on these scales which would

[2] While the use of a criterion group of only 25 (5.7%) means that the statistical measures obtained were less reliable than with a larger group, the degree of relationship indicated by correlation coefficients is not affected by such an extreme dichotomy as long as the distribution of both variables approaches normality. Furthermore, the criterion itself, since it implied the ability of teachers to recognize maladjustment and their decision to refer students to the college psychiatrist, was considered to be much more reliable when limited to extreme cases than if cases of moderate maladjustment had been included.

TABLE 1

Relationship of Bernreuter Inventory Scores to Serious Maladjustment as Judged by a Psychiatrist (N = 438)

Bernreuter scale	Coefficient of biserial correlation (r_{bis})[3]	Standard error of r_{bis}
B1–N (Neuroticism)	+.40	±.09
B2–S (Self-sufficiency)	+.28	±.09
B3–I (Introversion)	+.45	±.09
B4–D (Tendency to dominate)	−.18	±.10
F1–C (Lack of self-confidence)	+.29	±.09
F2–S (Unsociability)	+.50	±.08

[3] Not corrected for attenuation. Such correction would increase these coefficients by 25 per cent or so.

readily serve to pick out the majority of the maladjusted students without at the same time drawing attention to too many of the others. A score falling in Bernreuter's 80th

TABLE 2

Incidence of "Serious Maladjustment" Among Students with High Scores on the Bernreuter Personality Inventory

Scale	Students with high Inventory scores	"Maladjusted" students with high scores	Incidence of "Maladjustment"	P[4]
F2–S	84	14	17%	.00003
B1–N	54	10	19%	.0003
B3–I	33	7	21%	.0013
F1–C	47	8	17%	.0025
B2–S	98	11	11%	.011
B4–D	37	5	14%	.051

[4] "P" is the probability that a number of maladjusted students, as great as or greater than that observed in each group, would occur through the operation of chance factors alone. P is thus a second indication that there is a statistically reliable relationship between high Bernreuter scores and serious maladjustment. Where the value of P is less than .01, the relationship indicated is considered to be highly reliable.

percentile or above on five of the scales, and in his 19th percentile or below on the B4–D scale, was used to indicate possible maladjustment. Table 2 shows the efficiency of this method of prediction.

This table demonstrates that high Bernreuter Inventory scores can be used to predict the incidence of serious maladjustment to a reliably greater degree than is possible with no instrument of prediction whatever. In the whole student group, the incidence of serious maladjustment was only 6 per cent. In the groups selected by high scores on single scales of the Inventory, the incidence was from 11 per cent to 21 per cent. By using a single scale (F2–S), a majority of the 25 seriously maladjusted students could have been indicated to the college authorities before they began their college course.

Since there is a reliable correlation between maladjustment and scores on three of these scales, very high scores would be much more significant than many of the scores used in this study. If only the scores above the 90th percentile had been used, the incidence of maladjustment would have been far higher than that given in Table 2. This consideration would naturally be given weight in using this test in selecting students for admission, for example. However, limiting attention to those who had very high scores would mean that many cases of maladjustment would fail to be discovered by means of the test. In practice, the choice of a score to be considered significant depends upon one's preference for a high incidence or for a wide coverage. The use of the 80th percentile gives some weight to both considerations.

In an attempt to provide even greater coverage, without covering all the well-adjusted students too, the scales showing the highest relationship to maladjustment were combined in several ways. The results obtained with these combinations are presented in Table 3.

While these combinations made possible the coverage of 80 per cent of the seriously maladjusted students, the percentage

TABLE 3

Incidence of "Serious Maladjustment" Among Students with High Scores on the Bernreuter Personality Inventory

Either or both of the following scales	Students with high scores	Maladjusted Students with high scores	Incidence of "Maladjustment"	r_t[5] $\pm \sigma$	P
F2–S or B1–N	120	17	14%	$+.51 \pm .11$	.00002
F2–S or B2–S	123	17	14%	$+.51 \pm .11$	.00002
F2–S or B3–I	102	15	15%	$+.48 \pm .11$	.00005
B1–N or B3–I	55	10	18%	$+.46 \pm .15$	.0003
F2–S or B1–N or B2–S	159	20	13%	$+.56 \pm .10$	.00001
Any scale	173	20	12%	$+.52 \pm .10$	.00003

[5] Not corrected for attenuation.

incidence of maladjustment declined considerably from that obtaining in the case of most of the single scales. This was due to the fact that high scores on two or more scales were more frequent among the maladjusted group than among the others, so the inclusion of new scales did not add cases as rapidly to the former as to the latter.

It is apparent that the use of more than two, or at the most, three scales adds little to the predictive efficiency of the Inventory. And, since the scoring of these tests is very time-consuming, there seems little point in scoring more than scales B1–N, B2–S and F2–S, for practical purposes of prediction of maladjustment in college. High scores on any of these might, however, justify scoring the whole test, for the incidence of maladjustment was very pronounced where a high score on one scale was corroborated by a high score on another scale. Such cases are presented in Table 4.

It will be noted in this table that, in the case of scales which are known to have high inter-correlation coefficients (*e.g.*, F2–S and B2–S, or B1–N and F1–C), the incidence of maladjustment among students with high scores on both scales is no greater than among students with high scores on at

TABLE 4

Relationship of Serious Maladjustment to High Bernreuter Inventory Scores on Both of Two Scales, in Certain Combinations

Both of the following scales	Total students with high scores	"Maladjusted" students with high scores	Incidence of serious maladjustment	Observed multiple r_t[6]	Theoretical multiple R_t[7]
F2–S and B1–N	18	7	39%	+.64	.57
F2–S and B2–S	58	8	14%	+.34	.51
F2–S and B3–I	16	6	38%	+.61	.58
F2–S and F1–C	14	5	36%	+.57	.56
B1–N and B3–I	33	7	21%	+.46	.46
B1–N and B4–D (low)	26	5	19%	+.40	.46
B1–N and F1–C	44	8	18%	+.43	.44
F2–S and any other	71	12	17%	+.47	
B1–N and any other	50	10	20%	+.49	
B2–S and any other	59	8	14%	+.33	
Other combinations which were too infrequent for any reliable inferences:					
F2–S and B4–D (low)	10	3	30%		.54
B1–N and B2–S	6	3	50%		.58
B2–S and B3–I	5	2	40%		.59
B2–S and B4–D (low)	2	1	50%		.44
B2–S and F1–C	4	1	25%		.56

[6] Not corrected for attenuation.

[7] Not corrected for attenuation. These R_t's were calculated by the multiple correlation formula which employs the simple coefficients of correlation. Tetrachoric coefficients of correlation with the criterion were used, but the inter-correlation coefficients between the scales were obtained from this group of 438 students by the product-moment method.

least one of the scales (Table 2). But in the case of many of the other combinations, some of which occurred among very few students, the incidence of maladjustment was unusually high (*e.g.*, F2–S and B1–N, F2–S and F1–C). The size of the uncorrected actual and theoretical coefficients of multiple correlation indicates that, where such combinations of high scores were found, prediction of maladjustment can be made with a substantial degree of success.

It has frequently been suggested that family position is related in some significant way to personal adjustment. In an effort to test this hypothesis, the 25 seriously maladjusted students were classified by family position and compared with the numbers of all the students in each family position.[8] This comparison appears in Table 5.

In this table, there is found no indication that maladjustment is related to size of family, except an unreliably greater incidence in families with five or more siblings. Even this latter disappears when the relationship to family position is first parcelled out.

There is no indication of association of maladjustment with family position in the case of one-child or two-child families. In three-child families, however, there is some indication that maladjustment is associated with being the eldest or the youngest child, and in larger families this association is considerably more pronounced. The most significant association appears in the case of families with four or more children, where the intermediate children, numbering more than the eldest and youngest combined, account for only 1 out of 9 cases of serious maladjustment. When the incidence of maladjustment in the eldest and youngest positions combined is compared with that for all other children, the reliability of the association is indicated by the fact that an association as

[8] The 438 students in this study did not include twins where either was the oldest or the youngest in a family ($N = 8$), or cases where a sibling was known to have died ($N = 2$), since the family position in such cases was not strictly comparable to the positions used in the classification of Tables 5 and 6.

TABLE 5

Relationship of Position in Family to Serious Maladjustment in College. (There Were 25 Seriously Maladjusted Students, or 5.7 Per Cent of the 438 Studied)

Size of family and position in it	Total number of students	Number of seriously maladjusted students	Incidence of serious maladjustment	
			Per cent	
Only children	73	3	4	
Families with 2 children				
Eldest child	73	3	4	
Youngest child	68	4	6	
Total	141	7	5	
Families with 3 children				
Eldest child	42	3	7	
Middle child	37	1	3	
Youngest child	39	2	5	
Total	118	6	5	
Families with 4 children				
Eldest child	17	2	12	
Intermediate children	34	0	0	
Youngest child	14	2	14	
Total	65	4	6	
Families with 5 or more children				
Eldest child	10	3	30	
Intermediate children	23	1	4	
Youngest child	8	1	13	
Total	41	5	12	
Total of all children	438	25	5.7	
Families with *4 or more* children				P[9]
Eldest child	27	5	19	
Intermediate children	57	1	2	
Youngest child	22	3	14	
Total	106	9	9	
Eldest or youngest	49	8	16	.0033
Families with *3 or more* children				
Eldest child	69	8	12	
Intermediate children	94	2	2	
Youngest child	61	5	8	
Total	224	15	7	
Eldest or youngest	130	13	10	.013

[9] In Tables 5 and 6, "P" represents the probability that a number of maladjusted students, as great as or greater than that observed in each group, would occur through the operation of chance factors alone.

marked as this would arise by chance only 3 times in a thousand.

The addition of three-child families to the larger families reduces the significance of family position somewhat, but the association of maladjustment with being the oldest or the youngest is still marked enough so that it would occur by chance only 1 time in a hundred. And the incidence in these two positions together (10 per cent) is much higher than the incidence (3.9 per cent) for all other positions combined.

The significance of family position as an indicator of possible serious maladjustment is much increased when it is combined with high Bernreuter Inventory scores. This is shown in Table 6, where the 120 students with scores in the top two deciles on scales F2–S or B1–N are classified with regard to family position.[10]

Again, it is evident that the eldest and youngest children in families of three or more children were in an unfavorable position, particularly if they also had high Bernreuter Inventory scores. The incidence of serious maladjustment among this group was 24 per cent, or nearly seven times as great as the incidence (3.6 per cent) among the children in more favorable family positions. While the method of tetrachoric correlation is not strictly applicable to a table of this sort,[11] the extent of relationship between maladjustment and a combination of a high-Inventory-score-with-an-unfavorable-family-position can be suggested by the calculation of such a coefficient. In this case, its value was +.57. Correction for attenuation would raise this above +.70. That some relationship is present in this particular combination of variables is established by the fact that a table as extreme as this would arise by chance less than once in a hundred thousand times.

[10] B1–N was chosen for this purpose instead of B3–I because the coverage of F2–S and B1–N was somewhat greater than that of F2–S and B3–I (see Table 3).

[11] It is not applicable because the division into "favorable" and "unfavorable" family positions does not resemble the division of a continuous normal distribution into two parts.

TABLE 6

Relationship of Position in Family to Serious Maladjustment in College, Among 120 Students, All of Whom Had F2–S or B1–N Scores in the Highest Two Deciles

Size of family and position in it	Total number of students	Number of seriously maladjusted students	Incidence of serious maladjustment	
			Per cent	
Only children	24	2	8	
Families with 2 children				
Eldest child	21	2	10	
Youngest child	10	1	10	
Total	31	3	10	
Families with 3 children				
Eldest child	15	3	20	
Middle child	6	0	0	
Youngest child	10	1	10	
Total	31	4	13	
Families with *4 or more* children				
Eldest child	13	4	31	
Intermediate children	14	1	7	
Youngest child	7	3	43	
Total	34	8	24	P
Eldest or youngest	20	7	35	.00004
Families with *3 or more* children				
Eldest child	28	7	25	
Intermediate children	20	1	5	
Youngest child	17	4	24	
Total	65	12	18	
Eldest or youngest	45	11	24	.000006
Total of all children	120	17	14	

In every case where this study demonstrates that certain factors are related to maladjustment, it should be remembered that the only way that maladjustment was discovered here was through its recognition by the teachers themselves, followed by interviews with the psychiatrist. The success of prediction was thus measured by a criterion which was unreliable, in the sense that some cases of maladjustment

undoubtedly escaped notice at the level of observation by teachers. Anything that would aid these teachers or faculty advisers in studying the development of students would be expected to increase the reliability of this criterion by making it less likely that maladjustment would be overlooked. And a more reliable criterion would clearly be accompanied by a higher degree of correlation between the criterion and the predictive factors.

In other words, if faculty advisers used such information as Bernreuter Inventory scores and family position as significant clues to follow up in the study of their students, this in itself would be expected to increase the statistical reliability of the method. It should be emphasized, in this connection, that the final criterion employed in this study was the judgment of the psychiatrist and, while the Inventory scores were available to him, his judgments were based on overt behavior, rather than on the Inventory scores.

CONCLUSIONS

1. There is a significant and reliable relationship between high scores on the Bernreuter Personality Inventory (taken on entrance to college) and serious maladjustment uncovered later in college. Scales F2–S, B1–N and B2–S are most important in this connection. Scoring of the other scales adds to the effectiveness of the test in predicting possible maladjustment only when they are used to corroborate a high score on one of these three.

2. There is a significant and reliable relationship between serious maladjustment in college and being the eldest or the youngest child in families with three or more children.

3. Knowledge of Bernreuter Inventory scores and family position can be used to select, at the time of college matriculation, a group of students among whom the incidence of serious maladjustment in college will turn out to be much greater than among the student body as a whole. The use of this method of prediction would in itself tend to increase the method's effectiveness.

THE MEASUREMENT OF TRADE NAME FAMILIARITY

EDWIN E. GHISELLI
University of California

TRADE name familiarity is typically measured by the pure recall method. In the laboratory or in class room demonstrations the subjects are given a list of commodity names and are instructed to write down the first brand that comes to mind for each commodity or to write down all of the brands they can recall of each commodity. In market research field investigations respondents are usually asked questions such as "What brands of commodity X do you know?"

When instructions are given to recall all of the brands possible, either the per cent of people mentioning a brand *first* or the per cent of people mentioning it in *any serial position* may be taken as an index of familiarity. The first measure may be termed the per cent of first mentions and the second the per cent of total mentions. The most extensive studies of trade name familiarity using this pure recall method have been made by Hotchkiss and Franken.[1] In the analysis of their data these investigators usually used the per cent of first mentions although in their later studies they obtained the per cent of total mentions. It remains to be seen which of these indices is the better measure of familiarity.

The argument has been advanced that the first association is the strongest and hence the per cent of first mentions is the better index of familiarity. On the other hand it has often been asserted that the order of responses is largely a matter of chance. There does seem to be some justification for holding that the first response is not necessarily the first trade name recalled. A number of subjects have voluntarily reported to the writer that they did not respond with the first name that

[1] G. B. Hotchkiss and R. B. Franken, *The Leadership of Advertised Brands*, 1923; *The Measurement of Advertising Effects*, 1927.

came to mind because they were certain they knew it and could write it down at any time. They preferred, rather, to concentrate on recalling names which were not so easily called up. However, if the brands were reported in a purely chance order one would not expect to find a given brand reported first with any such frequencies as those found by Hotchkiss and Franken to hold for some commodities.

Instead of attempting to justify purely on logical grounds the use of one index of familiarity rather than the other it would seem to be more profitable to determine whether the two measures give the same or different results. The writer has computed the coefficients of correlation (correlation ratio) between the per cent of first mentions and the per cent of total mentions from Hotchkiss and Franken's data. These coefficients are given in the first column of figures in Table 1. With a group of 120 undergraduate students the writer also tested the familiarity of six of the commodities studied by these investigators. The coefficients of correlation between the two measures of familiarity obtained in this study are given in the second column of figures in the table. All but three of the coefficients are .90 or higher. This indicates that for most commodities the rank order of the brands in terms

TABLE 1

Coefficients of Correlation between Per Cent of First Mentions and Per Cent of Total Mentions

	Hotchkiss and Franken	Ghiselli
Dentifrice	.95	.92
Soap	.90	.....
Hosiery	.94	.....
Typewriters	.91	.....
Breakfast Food	.80	.72
Coffee	.97	.97
Cigarettes	.90	.78
Fountain Pens	.98	.99
Hats	.98	.....
Watches	.95	.99

of the per cent of first mentions corresponds closely to the rank order in terms of the per cent of total mentions.

It should be pointed out that the relationships are not linear,

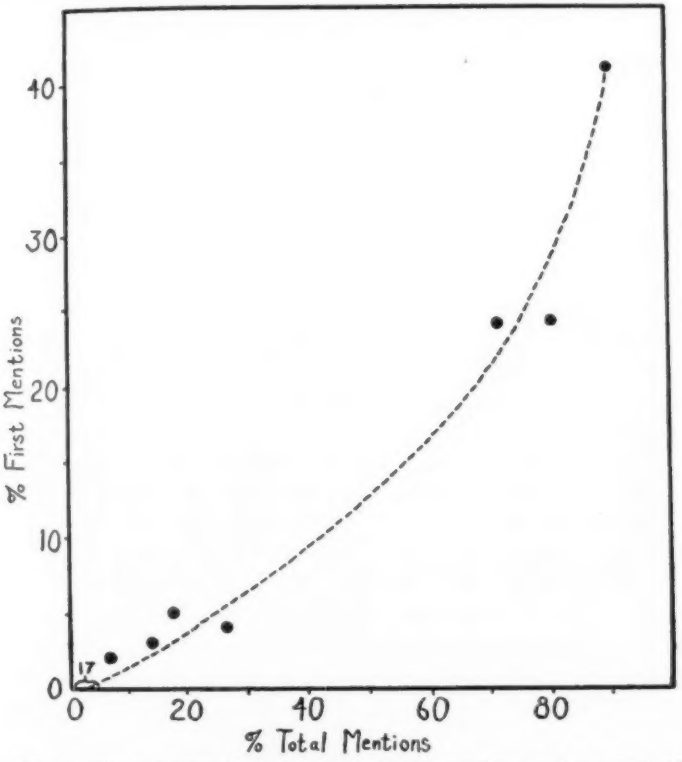

FIG. 1. The relationship of per cent of first mentions and per cent of total mentions for fountain pens.

the degree of curvature of the regression lines varying for different commodities. Hence the coefficients of correlation given in Table 1 are the means of the two correlation ratios (the regression of x on y and the regression of y on x) computed from each scatter. The scatter diagram for the data obtained for fountain pens, a fairly typical case, is shown in Figure 1. It should be noted that as trade name familiarity

increases in terms of per cent of total mentions there is a disproportionately greater increase in terms of the per cent of first mentions.

The frequency distributions yielded by the two indices of familiarity for the twenty-four brands of fountain pens which were mentioned are given in Figure 2. If the per cent of first

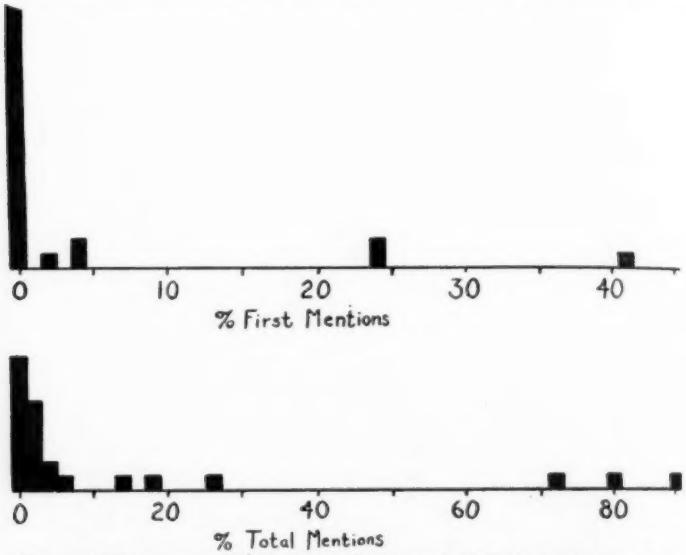

FIG. 2. Frequency distributions of per cent of first mentions and per cent of total mentions for fountain pens.

mentions were taken as the index of familiarity it would be concluded that with fountain pens only one brand is very outstanding, two being fairly familiar, and the large majority relatively unfamiliar. However, if the per cent of total mentions were used it would be concluded that there are three outstanding brands.

It is clear, then, that although the two indices of familiarity tend to be highly correlated they none the less give different pictures of the familiarity of trade names in the minds of the public. However, at the present time no conclusion can be drawn as to which is the better index.

STUDY HABIT INVENTORY SCORES AND SCHOLARSHIP

H. PHOEBE GORDON
The Psychological Corporation

INTRODUCTION

GOOD study habits have long been assumed to be distinct assets in academic achievement. Since this seems apt to be particularly true at the collegiate level, several study habit inventories have been developed in an attempt to obtain an objective measurement of the study skills of college students. Dr. C. Gilbert Wrenn and Dr. Norman M. Locke have both developed such inventories.[1] They have used as a criterion the ability of the inventory scores to differentiate between high and low scholarship groups in college courses. Since both authors present statistical evidence that their inventories make this differentiation, they may quite possibly be justified in feeling that the items in their inventories possess "considerable diagnostic value in portraying the study habits that influence a student toward high or low scholarship."[2]

These authors also appear to feel that, as a result of their findings, the inventories may be used for students about to enter college courses, in order to predict the assets or handicaps which they will find in their study skills. As Dr. Wrenn puts it, the inventory may be used "with an entering class of college freshmen or a college preparatory class of high school seniors, in order to determine the habits of work that will prove an immediate handicap to scholastic adjustments."[3] Locke does not specifically mention entering students but seems to

[1] C. Gilbert Wrenn, *Study Habits Inventory*, Stanford Univ. Press, and Norman M. Locke, *Student Skills Inventory*.

[2] C. Gilbert Wrenn, Manual to *Study Habits Inventory*, page 2.

[3] *Ibid.*

imply that they might well be given the inventory when he states that "students can be classified as to study habits on the basis either of total score or subtest score" and that "these classifications can be used for guidance programs or orientation courses."[4]

It is with this predictive use of the study habit inventories that the present writer takes exception, not on the grounds that the items used necessarily fail to differentiate between good and bad study habits, but on the grounds that the average high school senior who has done acceptable or perhaps even outstanding work in his high school courses, is often neither able nor willing to make a correct analysis of his own study skills in relation to work of collegiate level.

Take, for instance, note-taking techniques. How many students in the average high school have been subjected to the lecture type of instruction and have had opportunity to develop independent note-taking techniques? And again, take the student who is in the upper fourth of her high school group in scholastic aptitude and who gets on the "honor roll" without spending any more time in studying than is provided by the "study periods" assigned to her during school hours. Does she know anything of planning her own time for study or arranging efficient study situations? With these and similar questions in mind it seems essential to know something of the scores in study habit inventories made by students about to begin their college work in relation to *subsequent* achievement.

The material published by Dr. Locke and Dr. Wrenn does not give any direct evidence regarding the predictive value of study-habit questionnaires. So far as can be determined both inventories have been studied only in relation to students already in colleges. These students had already been exposed to academic work of college level and knew for themselves what scholastic record they were making. It seems quite pos-

[4] Norman M. Locke, "A Study Habits Test," *J. Appl. Psychol.*, Aug., 1940.

sible that a student who has already been faced with a low record in his class work is going to be more aware of his study habit limitations and more ready to admit them than the entering student with high hopes of the record he will achieve and with no knowledge of the difficulties he will face. In that case it is quite conceivable that a student answering the inventory before he has faced these problems or found any inadequacies in his techniques would answer questions concerning his study habits quite differently from the way he would answer the same questions after experience with class work at the collegiate level.

In an effort to acquire information on this phase of the problem, the following investigation was made.

METHOD AND DATA USED

The study habits scores of 118 high school graduates who entered Bellevue and Kings County Schools of Nursing[5] in the fall of 1936 were obtained. For this purpose the Wrenn Study Habits Inventory was used. The inventory was given during the orientation week for the two schools, before any class work was undertaken. The same students had previously taken the Alpha Examination (Wells Revision, Form V) so that a measure of scholastic aptitude was available. The Alpha scores ranged from 108 to 203. As an index of scholarship the students' grades in the class work of the first semester were used. For this purpose only their academic courses, such as Anatomy, Physiology, etc., were used, since these might be expected to be most nearly comparable to college courses. The average of the grades in these courses for the students in the group ranged from 74.9 to 94.6.

Comparing the grades, Alpha scores, and Study Inventory scores for these 118 students, the following coefficients of correlation were obtained:

[5] These schools are two of the schools of nursing of the Department of Hospitals of New York City.

TABLE 1

1. Scientific Course Grades
2. Alpha Scores
3. Study Inventory Scores

r_{12}	.582
r_{13}	.107
$r_{12.3}$	.093

It would seem reasonable to assume that if study habit scores made by students about to enter college are as valid as those made after entrance, the mean scores of high and low scholarship groups taken from this entering group would show the same type of differentiation that Dr. Wrenn's figures show for his groups of college students. According to the findings of Dr. Wrenn, the median for the high scholarship group was 46 and for the low scholarship group was – 12.[6] This differs markedly from the following figures:

TABLE 2

Mean Scores in Study Habits Inventory in Respect to Placement by Science Grades

Quartile	Mean Score
1 (lowest)	38.0
2	51.4
3	43.9
4 (highest)	43.4
Total	44.3 (N = 118)

It will be noticed that the lowest quartile in Science grades has a mean score only 6 points below the mean score of the highest quartile. Since standard deviation for the entire group is 33.6 this difference is negligible. It will be noticed that the mean for the entire group of 118 is 44.3, almost identical with that of Dr. Wrenn's high scholarship group. Yet there is no reason to believe that these students, some of whom received very low grades in their nursing school courses, would all have achieved high grades if they had undertaken college work.

[6] C. Gilbert Wrenn, Manual to *Study Habits Inventory*, page 3.

In an attempt to check further on the relationship between study inventory scores for entering students and subsequent scholastic achievement the Wrenn Inventory was given to the entering class of 27 student nurses at another hospital in New York City. It was administered by a psychologist who was known to the students to have no connection with the school and who made it clear that the resulting scores would not be reported to the school. This was done to remove any temptation for the students to try to answer the questions simply to win approval of the instructors. At the end of the first month, after the students had been exposed to the classwork in the school of nursing, the inventory was given again. For an index of scholastic achievement the grades in their major course of the first semester were used, namely Anatomy and Physiology. The range of grades in the course was from 68 to 89. The following coefficients of correlation resulted:

TABLE 3

1. Anatomy and Physiology Grades
2. Study Habit Inventory Scores at Entrance
3. Study Habit Inventory Scores at End of 1st Month
4. Alpha Scores

$r_{13} = .26$	$r_{14} = .43$
$r_{12} = .58$	$r_{24} = -.13$
$r_{23} = .72$	$r_{34} = .02$

What is perhaps even more important, since the Inventory is recommended for individual diagnosis, is that in 25 out of the 27 cases the score in the study habits inventory was lower at the retest than at the first trial. The mean scores on the two trials were as follows:

TABLE 4

Mean score for Study Habits Inventory at entrance	43.2
Mean score for Study Habits Inventory after one month in school of nursing	21.7
Difference	21.5
Standard error of Difference	8.37

SUMMARY AND CONCLUSIONS

1. The mean score of 44.3 in the Study Habits Inventory obtained for students entering Bellevue and Kings County Schools of Nursing and the mean score of 43.2 for students entering the other school of nursing correspond closely with the median score of 46 given by Dr. Wrenn for his "high scholarship" group who took the same inventory after having been in the University at least one semester. Since the subsequent record made by these 145 nursing school students indicates that they were not an outstanding group in scholastic achievement, it appears evident that the scores cannot be taken to predict academic success. Further, the mean score of the low scholarship quartile of the 118 students was almost identical with that of the high quartile for the same group, 38 and 44.3 respectively, both being close to the median of Dr. Wrenn's high scholarship group, while coefficients of correlation between study habit inventory scores and class grades for these entering classes were found to be .11 and .26. From these findings one would infer that the entering students were unable or unwilling to evaluate their study skills in a way that would differentiate low scholarship from high scholarship groups.

2. When grades in class work were correlated with study habits inventory scores made after the students had been attending nursing school classes for a month the coefficient proved to be .58 as compared with .26 when scores in the inventory were obtained from the same group of students before admission. Again the inference seems to be that entering students are not as able or willing to make accurate evaluation of their study habits as are students after they have had experience in classes of collegiate level. However, another interpretation of the figures suggests itself for consideration. It is possible that students are allowing their own evaluation of their study habits to be markedly affected by the grade of work they know they are doing. Thus, a poor student may decide that her note-taking techniques are poor and a good

student may consider hers good simply because they feel that their skill in these techniques in some way offers explanation of their good or bad grades. There is nothing in the present data nor in those presented by Dr. Wrenn or Dr. Locke to prove or disprove this possibility but it is a possible interpretation that has marked bearing on any value we may set on study scores and therefore needs further investigation.

3. Of the 27 students who took the Inventory at entrance and again one month later, 25 had a lower score at the time of the second test, the average decrease being 20.2 points. Since these students presumably did not have poorer study skills at the time of the second test than at the time of the first, the inference is that one cannot safely make use of the Study Habit scores of an entering student for individual diagnosis or classification.

Since Study Habit Inventories are of the self-rating type and have some similarity to personality or interest inventories, it is interesting to conjecture to what extent some of the above findings may be applicable to the latter type of inventories as well. That is, may not an individual analyze her interests or attitudes quite differently in the light of success or failure in a given situation? If so, a student who has liked and been liked in the high school situation may give a picture of her personality quite different from her evaluation of herself after she has experienced the contacts and adjustments of the new situation. Further investigation along this line seems indicated before we can rely on the predictive value of inventories of this type.

THE TIMBRE TEST IN THE REVISED SEASHORE MEASURES

DON LEWIS
State University of Iowa

IN the revised edition of the Seashore Measures of Musical Talents, which was recently issued by the RCA Manufacturing Company, a timbre test appears in place of the old consonance test.[1] This new test was designed to measure capacity to discriminate between complex sounds which differ only in harmonic structure. It is made up of two series (A and B). Each series consists of fifty pairs of complex tones, and each one is recorded separately on a single face of a 12-inch record. Half of the pairs in each series are the same in structure, the other half different. An observer is instructed to judge whether the two tones of a pair are the same or different in timbre; and his score on the test is determined by the number of correct responses he makes. Norms are available for both school children and adults.

The various stimulus tones for the test were produced with a special generator. All of them are constituted of a fundamental component and its first five overtones. The fundamental frequency is 180 cycles. Changes in structure are due to reciprocal alterations in the intensities of the 3rd and 4th harmonics.

The first step, in the actual construction of the test, was the selection, after successive trials, of a musically-acceptable standard tone which could be paired with any one of several

[1] Of the six tests in the original Seashore battery (pitch, intensity, time, rhythm, tonal memory, and consonance), all have been retained except the consonance test. Each one has been modified to a considerable extent but is basically unchanged. A description of the revised tests, together with norms and information on reliability, will be published at an early date in a monograph by Saetveit, Lewis, and Seashore.

properly-chosen variable tones. The detailed structure of the standard tone is depicted in Figure 1. In the figure, relative intensity is plotted against frequency. The number of each harmonic is also indicated along the abscissa. The solid vertical bars show the relative intensities of the six harmonics composing the standard tone. The intensity of the 3rd harmonic in the standard tone was arbitrarily chosen as a reference point and was assigned a value of zero db. Consequently, the intensities of the other harmonics, all below the reference point, are designated by negative db values. For example, the first harmonic, as shown in Figure 1, is 8 db below the level of the 3rd, the second harmonic 6 db below, and so on. The 4th, 5th, and 6th harmonics have relative intensities of −11, −12, and −14.5 db, respectively.

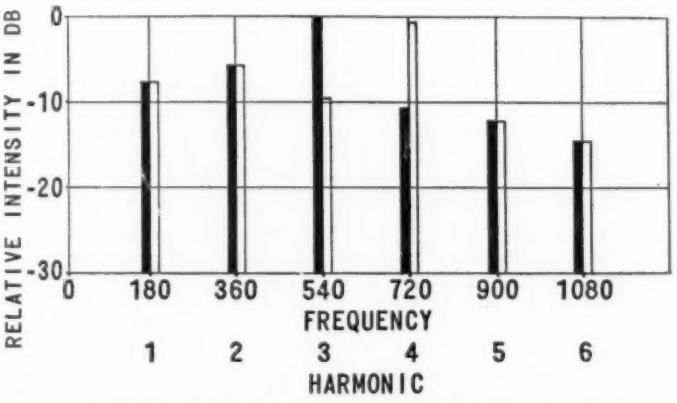

FIG. 1. Relative intensities (in db) of the six harmonics in the standard tone (solid bars) and in the particular variable tone (open bars) which differs most in structure from the standard. The values apply, of course, only to what is on the phonograph records, and not necessarily to the tones as actually produced by a given phonograph in a given room. (See the text for explanation.) The fundamental frequency of all tones in the test is 180 cycles. When, in keeping with instructions, the test is given at a "medium-loud" level, the 3rd harmonic in the standard tone has an intensity level (db above 10^{-16} watt) somewhere around 70 db.

There are ten different variable tones in the test–five in each series. The particular variable which differs most in structure from the standard tone is represented graphically by the open vertical bars in Figure 1. In this case, the intensity of the 4th harmonic is 10 db above its level in the standard. To offset the increase in the 4th, it was necessary to reduce the level of the 3rd harmonic by 9.6 db. This amount of structural difference between variable and standard is sufficient to give rise to a rather easily perceived difference in timbre. Smaller structural differences, of course, produce smaller perceptual differences. The two series in the test are each so constructed as to be increasingly difficult from beginning to end.

In each of the variable tones used, the 4th harmonic is a required number of db above its level in the standard tone, and the 3rd harmonic is below its level in the standard by an amount sufficient to keep total intensity unchanged. The intensities of the 1st, 2nd, 5th, and 6th harmonics remain fixed. Table 1 gives the various amounts by which the two altered harmonics differ from their levels in the standard tone. The variable tone portrayed in Figure 1 occurs in the first 10 stimulus pairs of Series A. The smallest structural difference employed occurs in the last 10 stimulus pairs of Series B, wherein the 4th harmonic is 3 db above its level, and the 3rd is 0.5 db below its level, in the standard tone. This difference is small enough to place a tax on the most superior of ears.

TABLE 1

Amounts, in db, by Which the Intensities of the 3rd and 4th Harmonics in Variable Tones in the Timbre Test Differ from Their Levels in the Standard Tone

Pairs in series where used	Series A		Series B	
	Increase in 4th harmonic	Decrease in 3rd harmonic	Increase in 4th harmonic	Decrease in 3rd harmonic
1–10	10.0	9.6	9.0	5.2
11–20	8.5	4.0	7.5	2.7
21–30	7.0	2.4	6.0	1.7
31–40	5.5	1.2	4.5	0.9
41–50	4.0	0.7	3.0	0.5

Series A is somewhat easier than Series B. It was specifically designed for use, along with the "A" forms of the other five tests in the Seashore battery, for general survey purposes. Series B, because of its greater difficulty, is recommended for use in testing individuals of superior capacity.

The timbre test, in its present form, is an outgrowth of studies of timbre that have been going on for several years in the Iowa laboratory. It is regarded as an improvement over a preliminary form which has already been described in considerable detail.[2] The test was included in the revised edition of the Seashore Measures only after considerable hesitation. There was some doubt as to whether or not a test of this type could be satisfactorily standardized for general use. Even now, doubt has not been completely dispelled. With modern electrical apparatus for sound production and control, it is possible to produce and to make phonograph recordings of complex tones of predetermined structure. The tones can be "put on" records in keeping with requirements. The important question is whether or not they can be "taken off" for test purposes, without excessive distortion. There are two principal complicating factors. One factor relates to the performance of typical phonographs. Any phonograph, regardless of its type, has its particular frequency-response characteristics, and the complex sounds that it will produce have structures which depend upon both the record being used and the characteristics of the instrument. The second disturbing factor is related to room acoustics. In a reverberant room, the position of an observer in the room will determine, to some extent, the type of stimulation he receives. Furthermore, the exact pattern of stimulation will be different from one room to another even though the sounds emitted by a loud speaker remain the same. It must be granted that if the actual stimuli constituting the timbre test cannot be kept fairly constant

[2] D. Lewis. "The Construction of a Timbre Test," *Psychol. Rec.*, 1939, 3, 115–136.

from one group to another or from one individual to another, an adequate standardization of the test cannot be made.

Preliminary surveys with the test have indicated that reasonably consistent results may be expected. The reliability of the test has been found to be rather satisfactory: it is as high as the reliability of the pitch test. The success that has been attained is probably attributable, in large part, to the precautions taken as a likely means of minimizing the operation of distorting factors. For one thing, frequencies are used in the stimulus tones which the average phonograph, be it electric or acoustic in type, will reproduce with a fair degree of fidelity, and which have sufficiently long wave-lengths to avoid alternate live and dead regions within a small area in a room. Furthermore, the beginning and ending of each tone is gradual rather than abrupt; so there are no important transients to disturb the structural pattern.

Further investigations need to be made to determine the maximum extent of variations in test results that can be expected for widely different acoustical conditions. However, if care is taken in the choice of the phonograph and in the placement of the instrument in the test room, and if the room is neither very small nor excessively reverberant,[3] the results obtained with the test should be sufficiently reliable for diagnostic and guidance purposes.

[3] Any modern phonograph is apparently suitable so long as its turntable runs smoothly. If there are "surges" in the turntable, the stimulus tones will not be steady and smooth, and judgments will be affected. The ordinary classroom is satisfactory for use unless its surfaces (ceiling, walls, and floor) are very hard and glazed, in which case reverberation is apt to be excessive. However, a room which is quite reverberant when it is empty becomes progressively less reverberant as the number of people in its increases. The phonograph should be placed six or eight feet from the nearest wall and about the same distance from the nearest listeners.

A TEST OF COMPETITIVE ADVERTISING[1]

ALFRED C. WELCH

Knox Reeves Advertising, Inc., Minneapolis

ESTABLISHING an attitude favorable to the sponsor's particular brand of a product is often the purpose of advertising. This attitude will be called BRAND PREFERENCE. Under conditions of *equivalent accessibility, equivalent price,* and *absence of outside inducements* such as contests and premiums, it seems legitimate to infer that the brand purchased is the brand preferred at the time. A test of competitive advertising, consequently, must either measure sales under these conditions or provide some other estimate to show how a given brand would rate if these conditions were attained.[2] Furthermore, it must provide an estimate of the contribution of advertising to Brand Preference independent of the influence of such factors as physical differences among brands and point-of-sale influences. Obviously, the Brand Preference status of a brand cannot be attributed to advertising if differences among brands in taste, appearance, etc., are great enough to be discriminated by a large proportion of consumers or if dealers influence the preference of consumers.

Cigarettes seem to meet these conditions under which the brand purchased should provide an accurate estimate of the brand preferred. All popular brands of cigarettes are equally

[1] Adapted from Chapter I of a thesis, "Consumer Attitude Criteria of Advertising Effectiveness," on file in the University of Minnesota Library. The writer gratefully acknowledges the help of Professor Donald G. Paterson.

[2] As early as 1932, Link realized the potentiality of Brand Preference as an important attitude in studying the effects of advertising and that Brand Preference might not correspond perfectly with sales under certain conditions (8).

accessible. In the sample of 966 University of Minnesota students tested during 1934–1937, most of the smokers smoked brands equal in price, and the preferences of non-smokers were concentrated among these same brands. The Old Gold Contest provided the only important outside inducement, and this contest comprised a special study.

Furthermore, it would be difficult to find a product less subject to point-of-sale influences than cigarettes; and three independent studies have concluded that smokers cannot discriminate differences among popular brands of non-menthol-ated cigarettes in most cases.[3] Since none of these studies was conclusive, however, the present study took the precaution of using non-smokers as a control.

Cigarettes, therefore, provide a special case in which the brands purchased should reflect accurately the effects of advertising. Consequently, brand purchases comprise a criterion against which a measure of Brand Preference can be checked.

THE BRAND PREFERENCE SCALE

As a measure of Brand Preference, the verbal expression of Brand Preference was used: each subject named five brands of cigarettes in order of preference. Arbitrary values of 100, 80, 60, 40, and 20 were assigned to the ranks. Non-mentions were treated as zeros. The mean of these values for each brand was used as an estimate of the Brand Preference rating of that brand.

While this simple treatment of Brand Preference responses represents a straightforward method of expressing Brand Preference as a single index, it is based upon a number of assumptions which must be investigated before the Brand Preference scale can be accepted.

The method assumes that Brand Preference ranks represent positions on a quantitative scale. To test this assumption 113 smokers were given the Brand Preference test twice, with a

[3] Husband and Godfrey (7); Watson, reported by Link (9); and Tinker and Welch (10).

week intervening. If the assumption is sound, shifts between adjacent ranks should be most frequent, shifts of two ranks should be second in frequency, etc. Table 1 shows that the frequency of shifting tended to decrease as the distance between ranks increased. Evidently, then, the assumption is sound.

TABLE 1

Comparison of the Frequency and Distance of Shifts in Brand Preference among 113 Smokers

Distance of shift	Mean number of shifts
1 Rank	20.5
2 Ranks	6.8
3 Ranks	2.8
4 Ranks	1.0
5 Ranks	1.0

The method also assumes that most consumers know at least five brands of cigarettes. Evidently, this assumption is sound because 99 per cent of 133 respondents named at least five brands when asked to name ten.

Selection of a mean of arbitrary values as the index of Brand Preference assumes that this method of estimating the Brand Preference status of each brand yields results that do not differ greatly from the results obtained by using other legitimate methods of expressing the Brand Preference status of each brand. According to Table 2, either frequency of first preference or number of mentions could be substituted for the Brand Preference index without changing the results seriously.

In this study the index was used because both of the other methods are likely to be less sensitive. Since previous use strongly conditions frequency of first preference and since a large proportion of consumers mention the leading brands of cigarettes, both of these methods are relatively insensitive. For commercial purposes either of these two methods might be preferred, depending upon the specific conditions of the study.

TABLE 2

Comparison of Brand Preference Index with Frequency of First Preference and with Number of Mentions As Ways of Expressing Brand Preference among 338 Smokers

Methods compared	Uncorrected r*	Degrees of freedom†	Corrected r‡
BP index and frequency of first preference	.912	27	.908
BP index and number of mentions	.977	27	.976

* $r = \dfrac{N\Sigma XY - \Sigma X \Sigma Y}{\sqrt{N\Sigma X^2 - (\Sigma X)^2}\ \sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$, in which X and Y are the raw scores of the two variables and N is the number of paired observations.

† In this analysis the brands mentioned, not the subjects, are the paired observations. Since N is only moderate and r is high, the standard error of r might be misleading (4, pp. 194–195). Consequently, the number of degrees of freedom is given so that one may use tables to test for significance (4, p. 212) or to find the probable minimum correlation (3, p. 393).

‡ Ezekiel's correction (3, p. 121). This correction provides a refined method of estimating r when the number of paired observations is small. In the above analysis, r and N are of sufficient magnitude to make the effect of the correction negligible.

For instance, Baker (1) found that the mean of arbitrary scale values was not a satisfactory method of expressing preferences for radio stations; and Gaudet (5) reported a rank-difference correlation of only .72 between frequency of first mention and number of mentions for favorite radio programs.

Finally, the arbitrary values used for the preference ranks rest on the very doubtful assumption that adjacent ranks are equidistant. Psychologically, it would be more reasonable to assume the greatest distance between the first and second preferences because brand habits in cigarettes are fairly rigid. From the practical viewpoint, the shifts between first and second preferences are most significant because they constitute gains and losses of sales. Since the assumption of equidistant ranks does not seem reasonable, an attempt was made to develop a scale on a more rational basis. If, during a given

week, a great many consumers change their fourth preferences to fifth preferences and vice versa while a smaller proportion change their second preferences to third preferences and vice versa, the distance between second and third must be greater than the distance between fourth and fifth ranks. The necessary data were provided by giving 113 smokers the Brand Preference test with a retest one week later. The scale separations for Scale I and Scale II, presented in Table 3, were based on the above reasoning. Scale II was derived from shifts between adjacent ranks only, while the method used for Scale I made use of all available direct and indirect estimates of each scale separation.[4] The scale separations for Scale III were the distances between the arbitrary values assigned to the ranks as explained previously.

TABLE 3

*Scale Separations of Three Brand Preference Scales**

Scale separations†	Scale I	Scale II	Scale III
First and second	65	53	20
Second and third	16	13	20
Third and fourth	9	11	20
Fourth and fifth	6	13	20
Fifth and not mentioned	4	10	20

* The calculations underlying the construction of these scales appear in the original thesis.

† The differences between adjacent ranks.

Obviously, the assumption of equidistant ranks was not confirmed. Nevertheless, from a practical viewpoint the use of Scale III can be justified if it will obtain the same results as the other two scales. With ease of computation in its favor, Scale III was accepted as the Brand Preference scale because Table 4 shows that it can be used without seriously changing the results obtained.

[4] For instance, the distance between third and fourth ranks (d_{34}) can be estimated indirectly by each of the following comparisons: $d_{13}-d_{12}$; $d_{24}-d_{23}$; $d_{35}-d_{45}$; $d_{3A}-d_{4A}$. (A refers to the "not mentioned" category).

TABLE 4

Comparison of Three Scales As Methods of Measuring the Brand Preference of 338 Smokers

Scales	Uncorrected r*	Degrees of freedom*	Corrected r*
I and II	.994	27	.993
II and III	.995	27	.994
I and III	.977	27	.976

* See footnotes of Table 2.

RELIABILITY OF THE BRAND PREFERENCE SCALE

The retest reliability was estimated by giving the test twice with a week intervening. The tendency for consumers to rank the same brands in the same positions is shown in Table 5.

TABLE 5

Retest Reliability of Brand Preference among 113 Smokers and 127 Non-Smokers

Group	Chi-square	Degrees of freedom	P	mC*
Smokers	689.7	16	less than .01	.81
Non-smokers	594.7	16	less than .01	.74

* The corrected contingency coefficients are probably the best estimate of r that can be made with the data available (6, pp. 356–359).

Although the consistency of the scale in measuring the Brand Preference of individual consumers was not entirely satisfactory, its practical value lies in its use in measuring the status of different brands. Its retest reliability with brands as the paired observations, consequently, is more critical. When brands were the paired observations, shifts to and from each brand tended to cancel each other; and the retest reliability was much higher. Among the same 113 smokers, for example, the retest reliability with brands as the paired observations was .998.[5] Of course it cannot be denied

[5] Corrected r (3, p. 121) with 24 degrees of freedom.

that this coefficient may be spuriously high because some responses made on the test may have been remembered when the respondents took the retest.

The sampling reliability of the Brand Preference index was estimated empirically by drawing at random pairs of samples from a homogeneous supply. The significance of the difference between the members of each of the 66 pairs was estimated by computing t (4, pp. 128–130). Then the empirical distribution of t was compared with the theoretical distribution. The two distributions paralleled each other closely enough for practical purposes; *e.g.*, 3 per cent of the obtained t's exceeded the .05 level of significance. Evidently the t-test provided a practical approximation of the proportion of times that random sampling yielded differences greater than specified magnitudes—although the assumption of a normally distributed underlying variate was doubtful. Consequently, a more exact determination of the sampling reliability of the Brand Preference index was not attempted.

VALIDITY OF THE BRAND PREFERENCE SCALE

For cigarettes, a valid test of Brand Preference should be related closely to brand purchases; because conditions of equivalent accessibility, equivalent price, and absence of outside inducements are approximated.[6] Actually 92 per cent of 323 smokers ranked first in preference the brands that they were smoking at the time,[7] and only 3 per cent ranked the brand used lower than second. This analysis, however, demonstrated the validity of the scale for the first preference only.

A critical test of the validity of the lower positions of the scale consists of its ability to predict the brands smokers will smoke if their preferred brands are not available. The test

[6] The Old Gold contestants constituted an exception to the condition of the absence of outside inducements. Therefore, they were excluded from the following two analyses.

[7] A conservative estimate of the percentage of respondents who reported accurately the brand being smoked was found to be 92 per cent by checking the responses against the brands being carried at the time.

was disguised as a study of the effect of breaking a habit. After taking the Brand Preference test, nineteen smokers agreed to give up their regular brands for one week and to smoke any other brands they wanted. The brands being smoked at the end of the week were found by actual check. The validity of the lower ranks of the Brand Preference scale was shown by the correlation of .87[8] between Brand Preference on the test (exclusive of the brand smoked regularly) and the number smoking each brand at the end of the week. This validity coefficient can be accepted as merely a rough approximation since only 11 degrees of freedom were available in computing the correlation. Furthermore, two conditions tended to give an underestimation of the true validity. First, only second and lower preferences could be included because the preferred brand of each smoker was ineligible as a brand to which he could switch. The validity of the scale as a whole probably was higher. Second, a larger group of smokers probably would yield a higher validity coefficient.[9]

The two analyses just presented established satisfactory validity for conditions under which brand purchases can be accepted as indicative of Brand Preference. If the Brand Preference scale is valid only under these conditions, it is subject to one of the limitations of the ordinary sales test; namely, the inability to measure the effects of advertising if an outside inducement is present. Since price differences among brands or changes in the price of the same brand constitute a special case of offering an outside inducement, any measure which resists the influence of these inducements will provide an especially valuable instrument for practical copy testing.

The Old Gold puzzle contest offered a special inducement to get smokers to buy Old Golds for a period of fifteen weeks.

[8] Corrected r (3, p. 121) with 11 degrees of freedom.

[9] This point may become clearer when it is remembered that the number of smokers in this analysis is analogous to the number of items in a test in analyses in which test scores are correlated.

During the contest 10.3 per cent of the 311 smokers tested were smoking Old Golds but only 6.1 per cent ranked Old Golds first in preference. The contestants provided the explanation for this discrepancy between use and preference: only 13 per cent of the contestants ranked Old Golds first whereas 92 per cent of the smokers of a brand usually ranked that brand first under normal conditions.[10] The question whether the verbal expression of preference or the brand purchased is the more accurate estimate of Brand Preference can be answered by finding which estimate can predict the post-contest status of Old Golds more accurately.

Consequently, a follow-up study was conducted to find what brands the contestants were smoking after the inducement had been withdrawn. To keep changes among non-contestants from obscuring the results, it was assumed that they were smoking the same brands as during the contest. Furthermore, the follow-up study was conducted before the winners had been announced, so that any emotional reaction of the losers would not affect the results. The only variable, then, was brand purchases among the contestants included in the follow-up study. Under these conditions, 6.2 per cent of 308 smokers were smoking Old Golds at the time of the follow-up study. Under conditions of outside inducements, evidently, Brand Preference was estimated more accurately by the verbal expression of preference than by actual use.

Further evidence that the mere smoking of a brand did not seriously affect its Brand Preference rating was provided by an analysis of changes in Brand Preference for brands which subjects smoked for one week merely in compliance with the investigator's request to change from the brand smoked regularly. Brand Preference increased only 9.5 points[11] and was 27 points below the index of brands smoked regularly.[12]

[10] In terms of the Brand Preference scale, this difference was 38 points, with a critical ratio of 9.1.

[11] $t = 1.58$, P: greater than .1; Student's t (4, pp. 126–128) was used because it provides a more sensitive test of significance when the corresponding values of the two series are positively correlated.

[12] $t = 7.42$, P: less than .01.

PRACTICAL VALUE OF THE BRAND PREFERENCE SCALE

The Brand Preference scale was the result of an attempt to circumvent many of the objections usually raised against the sales test and still heed the timely warning of Coutant, "A copy test that stops short of some form of sales test leaves wide open the question whether any of the advertising will sell merchandise" (2).

As a practical tool for measuring the effects of advertising, the superiority of the Brand Preference scale over the sales test is suggested by the following arguments:

1. A representative sample can be obtained at a relatively low cost because the test is based upon consumer interviews.

2. The Brand Preference index is more resistant to irrelevant influences because weather and similar factors can increase or decrease the number of units sold without seriously affecting the *relative* positions of brands.

3. Contests, premium offers, and other special inducements which may cause a temporary increase in sales have little effect on the Brand Preference index even if these inducements succeed in getting consumers to use the brand temporarily.

4. The sales test is a relatively insensitive instrument because many brand habits are rigid. The Brand Preference scale reduces the rigidity introduced by the habitual use of a particular brand by including five preferences. Furthermore, the sensitivity of the scale was suggested by a special study in which 84 subjects were exposed to a fountain pen advertisement after taking the Brand Preference test. The following day the index for this brand was significantly higher.[13] Since fountain pens are purchased infrequently, only an extensive sales test could detect the effects of this advertisement.

Yet the evidence submitted previously indicated that the Brand Preference scale was sound methodologically, that its

[13] Student's $t = 3.3$, P: less than .01. The results for four other advertisements were inconclusive. Three of them increased more in the experimental group than in the control group, but the increase was not significant statistically.

reliability was satisfactory, and that it was a valid test of the effects of competitive advertising if administered properly.

SUMMARY

As a test of competitive advertising, a Brand Preference scale was developed. The assumptions underlying the construction and use of the scale were tested. The retest reliability and the sampling reliability were determined. The validity of the scale was tested by finding the relationship between first preferences and actual sales; by predicting the brands to which consumers would switch when their regular brands were not available; and by predicting the post-contest sales of Old Gold cigarettes.

The resistance of the Brand Preference scale to irrelevant influences (*i.e.*, contests and other special inducements) that affect sales tests was demonstrated. Evidence was submitted that suggested that the scale is more sensitive than a sales test in detecting the results of advertising.

REFERENCES

1. Baker, K. H. Radio listening and socio-economic status. *Psychol. Rec.*, 1937, 1, No. 9, 110–111.
2. Coutant, Frank. Testing the marksmanship of advertising copy. *Adv. & Sell.*, July 15, 1937, p. 21.
3. Ezekiel, Mordecai. Methods of correlation analysis. John Wiley & Sons, 1930.
4. Fisher, R. A. Statistical methods for research workers. London: Oliver and Boyd, 1936 (Sixth Edition).
5. Gaudet, Hazel. The favorite radio program. *J. Appl. Psychol.*, 1939, 23, 118.
6. Guilford, J. P. Psychometric methods. McGraw-Hill, 1936.
7. Husband, R. W., and Godfrey, Jane. An experimental study of cigarette identification. *J. Appl. Psychol.*, 1934, 18, 220–223.
8. Link, H. C. The new psychology of selling and advertising. Macmillan, 1932, p. 45.
9. Link, H. C. Significance of change in cigarette preference. *Adv. & Sell.*, June 6, 1935.
10. Tinker, M. A., and Welch, A. C. A study of brand preference in cigarettes. 1934 (unpublished).

A CORRECTION

ALBERT B. BLANKENSHIP
New York City

RECENTLY the writer reported[1] a field study which indicated that the wording of the question was an influence upon the results obtained from a public opinion questionnaire. Following publication of this study, Dr. G. K. Bennett, of the Psychological Corporation, discovered two errors in the statistical analysis. One of these errors was of a major sort, while the other may be regarded as minor.

It was concluded in this study that the accuracy with which an election can be predicted is dependent upon the question wording employed. With this general finding there can be no argument. But consider one of the conclusions based upon the findings of the study (*Psychol. Rec.*, p. 398):

> "The objective variable question was found to be the most valid; it was reliably different from all of the other forms. Next most valid was the negative objective, which was reliably different from two of the other four question forms. The P.O.C.L., the positive objective and the subjective followed in that order; however, none of the last three was reliably different from one another."

These statements are based upon an analysis of the statistical reliability of the differences between the "predictive accuracy"[2] of the five wordings used, by the method of the

[1] A. B. Blankenship, "The Influence of the Question Form upon the Response in a Public Opinion Poll," *Psychol. Rec.*, 1940, 3, 345–424. Also A. B. Blankenship, "Does the Question Form Influence Public Opinion Poll Results?" *J. Appl. Psychol.*, 1940, 24, 27–30.

[2] This index of validity was determined as follows (*J. Appl. Psychol.*, footnote 4 on page 29): "The figures . . . were secured by subtracting the total accuracy deviation . . . from 100%. For example . . . the election returns came out exactly 50% for and 50% opposed. . . . The objective question form showed 49.5% for, 50.5% against . . ., a total loss of accuracy of 1%, producing 99.0% validity."

standard error and critical ratio. Statisticians know that when a standard error of a proportion is computed (as was the case here), the proportion must represent a statistic of a known population. In the study reported, the "percentages of validity" were not this type of statistic, but were *indices* of validity. The 99 per cent validity of the objective question form is therefore not a proportion of the population. The procedure utilized for computing the validity is sound, but standard error may not be applied to these figures. One possible way to determine which question form best predicted the election returns would be to secure the difference of each wording result from the election returns, and such differences could then be analyzed in terms of the critical ratio. Table 1 illustrates how this method would be applied. (It should be pointed out that in the community where the study was conducted, the election returns were precisely 50.0 per cent for, and 50.0 per cent opposed to the issue.) Using the criterion

TABLE 1

Showing Differences of Each Question Form from the Election Returns, and Reliability of these Differences

	Objective	Subjective	Positive Objective	P.O.C.L.	Negative Objective
Per cent favorable	49.5	57.5	56.3	44.9	53.6
Standard error	2.9	2.8	2.9	2.8	2.9
Difference from election returns	.5	7.5	6.3	5.1	3.6
Critical ratio	.17	2.68	2.25	1.83	1.24
Probability*	57	100	99	96	89

* The probability expressed here is the number of chances in 100 that the deviation is not caused by chance. Any probability of 99.5 plus is shown as 100. In determining the critical ratios, election returns have been regarded as a parameter, not as a statistic.

of 95 chances in 100 (or higher) as denoting a true difference, Table 1 shows that only the objective and negative objective forms predict within a deviation expected by chance.

If the affirmative answers to each question form (as shown

in Table 37 of the *Record* article) are compared by the usual technique of computing standard errors and critical ratios,

TABLE 2

Showing the Differences and Reliabilities of the Differences for the Affirmative Answers to the Five Question Forms

Forms Compared	Difference (in %)	Sigma Diff.	CR	P
Objective-Subjective	8.0	4.1	1.95	97
Objective-Positive Objective	6.8	4.1	1.66	95
Objective-P.O.C.L.	4.6	4.1	1.12	86
Objective-Negative Objective	4.1	4.1	1.00	84
Subjective-Positive Objective	1.2	4.0	.30	62
Subjective-P.O.C.L.	13.6	4.0	3.40	100
Subjective-Negative Objective	3.9	4.1	.95	83
Positive Objective-P.O.C.L.	11.4	4.0	2.85	100
Positive Objective-Neg. Obj.	2.7	4.1	.66	74
P.O.C.L.-Negative Objective	8.7	4.1	2.12	98

the present Table 2 is the result. According to this method of analysis, five of the differences must be due to other than sampling factors, and when these are corrected (according to the method described later in the present paper) they are all still barely above the 95 per cent point of probability.

With the changes suggested from Tables 1 and 2, the conclusion reached (Conclusion ♯2, p. 398 of *Record* article) ought to read:

> "The objective and negative objective question forms both predict the election returns within a chance basis; the other forms predict with such a degree of inaccuracy that this deviation cannot be accounted for by chance. Further analysis shows that these two forms were not reliably different from one another in terms of the affirmative (or negative) answers secured."

At least two statements (pages 396 and 397 of the *Record* paper) need similar modification, in that the objective form is referred to as being the most valid of the five forms used.

Before this analysis of validity was presented in the *Record* article, the five groups of persons used on the study were compared, to show that they were essentially similar in terms of sampling characteristics. Tables 12 through 34 (with the exception of Table 33) present the percentages of persons in each of five groups who possess various characteristics such as age, income level, race, etc. The comparison is made by showing the greatest difference between the five groups, and indicating the chances that this is a reliable difference (in terms of the standard error and critical ratio). In practically all cases the critical ratio is less than 3.00, so it was concluded that not only were the extreme groups comparable, but those in between must be also. The critical reader might be disturbed to find that many of the probabilities are 90 or above. Actually, the probabilities of a real difference are not so great as shown.

In all of these cases the largest difference is being selected, and when one difference is selected from ten possible comparisons, the obtained critical ratio ought to be divided by 1.9.[3] If this be done, there is only one critical ratio that approaches the 5 per cent point of reliability (Table 34 in the *Record*—the "yes" answers to Q. 7). All of the groups compared in this study are considerably more alike than appeared from the earlier method of comparison.

[3] See David Segel. *Differential Diagnosis.* Baltimore: Warwick and York, 1934.

NEWS AND NOTES

The February 1939 issue of the *Journal of Applied Psychology* was devoted to Research in Radio, the special editor being Dr. Paul F. Lazarsfeld. Again, for December 1940 a special issue was devoted to Progress in Radio Research, again edited by Dr. Lazarsfeld, now Director of the Office of Radio Research, Columbia University, New York City. The December 1940 issue may be obtained for $2.00 a copy by addressing the Psychological Corporation, 522 Fifth Avenue, New York City. A year's subscription including this issue may be had at $6.00. The February 1939 and the December 1940 issues may be had for $3.00.

The Midwestern Psychological Association will hold its annual meeting at Ohio University, Athens, Ohio, from April 10 to 12, 1941. The officers of the Association are Dr. E. A. Culler, University of Rochester, Rochester, New York, President and Dr. Robert H. Seashore, Northwestern University, Evanston, Illinois, Secretary-Treasurer. A tentative examination of the abstracts of papers submitted to the Committee on Program would indicate that the response has been hearty and that the quality gives promise of being good. Those wishing to make reservations for rooms should communicate directly with Dr. J. R. Gentry, Department of Psychology, Ohio University, Athens, Ohio.

The Ohio Academy of Science will hold its annual meeting in Cleveland, Ohio, on May 8, 9, and 10, 1941. Abstracts of all papers in Psychology should be in the hands of the Vice-President, Paul E. Fields, Ohio Wesleyan University, Delaware, Ohio, not later than March 20, 1941. It has been suggested that one session of Section E, Psychology, be devoted to "original or new and interesting demonstrations of psychological facts and principles," and that members who wish to present unique demonstrations which have proven to be effective, notify the Vice-President. For additional information concerning this meeting, one should write to the Vice-President, at the above address.

Available upon request is a pamphlet entitled "The Civilian Conservation Corps" published by the American Youth Commission of the American Council on Education. This contains the recommendations of the Commission based upon the findings of a three-year investigation of the

social and educational aspects of the CCC camps. After this investigation had been completed, the Commission was invited to assist in the development of an eighteen month experimental training program in ten CCC camps in the Fifth Corps Area which ended June 30, 1940. This study provided the staff of the Commission close contact with every phase of CCC administration. Such topics as the following are discussed in the pamphlet: Forerunners of the CCC, Purpose of the CCC, Who Are the Enrollees, Selection of Enrollees, Experience in Camp, General Results of Camp Life, Effects on Citizenship, Negroes in the CCC, and Administrative Set-up of the CCC. The following conclusions were reached: that the CCC camps have benefited more than 2,000,000 youth and advanced the cause of conservation perhaps a generation; that possibly the time may come when the camps will no longer be needed to absorb unemployed youth in large numbers. Work camps have already demonstrated training values for particular types of youth, however, that indicate a permanently useful future for some type of work camp program as a public training institution.

The Psychologists' Club of San Francisco was established on October 25, 1940. Its membership is comprised of members of the American Psychological Association and its affiliated societies who are interested in applied psychology. Its purposes are: first, to discuss and clarify the professional problems of applied psychologists; second, to forward the public appreciation and utilization of a sound applied psychology; third, to exchange news and views on recent psychologic books, articles, tests, techniques, etc.; fourth, to afford convenient and time-saving opportunities for the widely scattered psychologists of the San Francisco Bay Region to share the problems and experiences, the knowledge and the thoughts that arise from their diverse psychologic interests and pursuits. The officers of the new Club are: Winfield Wickham, University of California, president; Dorothy Pencharz, Marin Junior College, vice-president; Fern McGrath, Alto Psychologic Center, secretary.

The 1940 volume of the Occupational Index, complete with cumulative subject, author, and title index, is now available. It contains 967 new references to literature on more than 300 different occupations. The price is $6.50. Further information may be obtained from Occupational Index, Inc., New York University, 81 Washington Square East, New York City.

The American Management Association Conference on Industrial Relations was held at The Palmer House, Chicago, Illinois, on February 12, 13 and 14, 1941. This Conference came at a time when American indus-

try is confronted with the gigantic task of quickly setting up the bulwarks that will preserve it and protect American society. The program was devoted to the special problems which are bound to appear in a rapid expansion of forces and facilities for defense production. Speakers were secured who were best qualified by practical experience in the fields of personnel and industrial relations. Dr. Carl I. Hovland, of the Psychology Department, Yale University, presented a paper entitled "Basic Principles of Learning and Training."

The American Youth Commission of the American Council on Education has recently had prepared three leaflets in connection with its radio series, "Youth Tells Its Story," in order to describe various examples of community activity in behalf of youth more fully than was possible during the broadcasts. *Finding the Facts about Youth* concerns surveys used to gather information about the youth problem in three community areas so that constructive action could be undertaken. Many communities have already made such surveys. Others are making them. This booklet reports how three communities—samples of the big city, medium-sized community, and a small village with a rural setting—surveyed their youth and what they learned about youth's needs today. *New Strength for America* presents the unique recreational project in Columbus, Indiana, in which an intelligent coordination of resources which already existed made it possible for the youth of the community to enjoy a program of healthful, wholesome recreation. The third leaflet, *Rallying Resources for Youth*, describes the community council development in Los Angeles County, California; Dowagiac, Michigan, and Greenville County, South Carolina.

One of the most recent statements published by the American Youth Commission, entitled *Next Steps in National Policy for Youth*, outlines a coordinated program for local and national action on problems of education, health, recreation and employment. These recommendations are based on five years of research, survey and demonstration programs conducted by staff members. This statement again reviews the entire field of youth problems in relation to national policy for youth. Consideration is given to the problems of adjustment which are met by the men in the new military camps. Attention should also be given to the young men who remain at home until called to service and to those who have been rejected from training for various reasons. Programs for health and physical fitness as well as programs for mental health and morale are urgently needed.

Copies of these leaflets may be obtained without charge by writing to the American Youth Commission, 744 Jackson Place, Washington, D. C.

BOOK REVIEWS

STOGDILL, EMILY L., AND HERNDON, A. *Objective Personality Study: A Workbook in Applied Mental Hygiene.* New York: Longmans, Green & Co., 1939.

COFFIN, J. H. *Psychology of Personality* (Student's Outline Series). New York: Longmans, Green & Co., 1940.

Many psychologists like to organize the mental hygiene course around a study of the personality of each student. This method tends to vitalize the course by utilizing the student's own experiences. Instructors employing it will appreciate the new workbook by Stogdill and Herndon. It is organized in twenty-four assignments subdivided into four units, titled respectively "Social Adjustments," "Developmental History," "Habits and Reactions" and "Summary and Plan Making." Each assignment calls for answers to fairly specific questions about the student's behavior or history. A bibliography, short and in general fairly well-chosen, is appended to each assignment.

The workbook does not (may not be intended to) cover the usual course in mental hygiene completely. No mention is made of such familiar phenomena as complexes, rationalization, identification, etc. Some assignments, such as that on frustration, are likely to be quite meaningless to the student until interpreted by the instructor. No objective test items are included, and no norms are given for such answers as might conceivably be treated statistically. No attempt is made to go into the tabooed aspects of personality development, although there are points at which the instructor can, if the total situation permits, direct the student to consider such processes.

This is not, then, a workbook in the usual sense of the term, covering course contents and providing objective test material. It is just an extended form of autobiographical questionnaire, to which essay answers are to be given, with readings suggested to help the student understand why this variety of experience is considered important. But, unless personal conferences are frequent, the workbook will not help the student understand himself. I think it will be useful for instructors who organize the mental hygiene course around a highly personal approach (a method of which I approve).

Coffin has prepared an outline on the psychology of personality in the familiar green-backed series. It is rather inferior to most of its com-

panion volumes. This field is not sufficiently crystallized to lend itself well to the outline treatment, and Coffin seems to have some bizarre notions about personality. It is not a little startling to read: "c. The psychique operates at the: i. *Organic* level as a kind of unconscious, organic intelligence or purposive, teleological or hormic principle. Note the behavior of the white blood cells in case of infection." (p. 36.) This is the sort of stimulus which sends physiologists into minor hysteria.

Another illustration, from the other extreme of the field covered, is the equation: "Persons plus processes plus patterns plus products, equals Society, which is group mind in historical flow." (p. 42.) With adequate definitions of terms, such a sentence might have meaning, but in this bare outline, it is likely to influence sociologists in the direction of neurotic behavior.

There are some parts of the outline, notably those following Gordon Allport's authoritative work, which are quite good. Unfortunately, most of them lack adequate bibliography (very few of the outlined assignments have any mention of readings) and would be almost meaningless to the student in this skeleton form. On the whole, I should recommend this book for instructors who may want handy outlines of certain fields covered, but it will not prove useful and may sometimes be misleading to students.

ROSS STAGNER,
Dartmouth College

MUNCIE, WENDELL. *Psychobiology and Psychiatry. A Textbook of Normal and Abnormal Human Behavior.* St. Louis: The C. V. Mosby Co., 1939. Pp. 739.

It is hardly proper for a clinical psychologist in writing a review of a textbook in psychiatry to compose an essay on the relations of his field with such coincidental interests. The subtitle of the present volume merely emphasizes the tenuousness of the boundary between psychology and psychiatry. The psychiatrist is writing psychology and in this instance does it reasonably well.

Adolf Meyer has of course been long recognized as America's foremost psychiatrist. This has been particularly true because of his constant efforts to base clinical psychiatry on a sounder basis than either the mysticism dear to the psychoanalyst, or the physiology of mind which inheres in most medical thinking. Unfortunately Dr. Meyer has not found it possible to write a systematic treatise himself. Of the several attempts made by his students and fellow workers, this volume by Dr. Muncie is by far the best. The book is divided into four parts: (1) psychobiology—the study of normal behavior; (2) abnormal behavior—pathology and psychiatry; (3) treatment; and (4) historical survey.

According to Dr. Muncie it was the American Psychological Association which proposed, in 1911, that psychology should be introduced into the medical curriculum. However, the good intentions of the Association were not realized because of the arguments as to what psychology was. In the dilemma Meyer proposed that after all the important thing was objective study of people behaving, keeping in mind always that the people were first of all biological organisms, and that they were behaving in certain stimulus situations and that their behavior in the past had an effect on their behavior in the present. To such a common-sense view of psychology he gave the name "psychobiology." In 1911 this was a view quite foreign to professional psychologists; today it is commonplace. However, Dr. Muncie, and other psychobiologists seem to be unaware, or refuse to admit, that the science of psychology has changed its emphasis. The chapters of the first part of this book differ very little from an elementary course in the psychology of personality, although there is almost no recognition of contributions psychologists have made to this field. The very important pedagogical methods introduced by Meyer at Johns Hopkins Medical School of orienting the study of personality around a self-analysis of the students' own behavioral history is the basis of organization of the material in these chapters. Part I might well serve as a textbook in methods in clinical psychology.

In the second part are discussed in detail the various kinds of abnormal behavior. Here there is strict avoidance of the usual Kraepelinian nosology. So-called abnormal behavior is to be viewed as behavior resulting from exactly the same sort of conditions as normal behavior; therefore, to speak of dementia praecox or mania as though they were disease entities or even clear-cut clinical pictures is misleading. Muncie uses the Meyerian terminology, of which there has been much criticism, but which is necessary if one is to think in terms of reaction-patterns and avoid nosological entities. The discussion of various reaction types is illustrated with excellent cases.

The third section on treatment is perhaps the poorest in the book. While everyone working with human behavior problems knows that specific therapeutic rules are impossible, these chapters are disappointingly general. The theory of treatment, special therapeutic aids, and therapeutic notes on special reaction types are all considered, and are valuable in a limited way. It is not at all clear just how the various suggestions might be, or have been used with actual cases. In spite of the availability of cases as shown in Part II, there is not a single illustrative case in Part III. It is to be hoped that in future revisions Dr. Muncie will offer case illustrations of the important techniques of therapy.

The 182 pages of Part IV comprise a tabular summary of the history of psychiatric thought on the major types of behavior deviations. Brief

abstracts and bibliographies of ancient, European, American and psychoanalytic writers on neurasthenia, anxiety states, hysteria, dementia praecox, etc., are given in parallel columns. This historical summary, although somewhat awkward to use, is invaluable for reference.

It is necessary for every clinical psychologist to be acquainted with the current psychiatric work. For many reasons the "Meyerian school" of psychiatry appears to be of especial significance. Therefore, it is not too great a claim to say that this volume should be in the working library of every psychological clinic.

C. M. LOUTTIT,
Indiana University

GRAY, T. CLARENCE, AND VOTAW, DAVID F. *Statistics Applied to Education and Psychology.* New York: The Ronald Press Company, 1939.

The authors have prepared in *Statistics Applied to Education and Psychology* a book quite appropriate for a text in a beginning course in statistics. As might be expected from the positions of the authors, the book's explanations and examples are concerned primarily with the viewpoint of one interested in education and educational psychology rather than in purely scientific psychology, though the value to the last group is not in any way lessened by such an emphasis.

If the book can be said to have any merit over other texts in current use, it is due to the clarity of explanation, the presentation of sample situations, and the inclusion of an abundance of, for the most part, well-chosen exercises at the end of each chapter. The book can probably be used in a semester course without significant abridgment or omission of any important parts.

The virtue of complete explanations is lessened occasionally by detailed clarification of some simple topic out of proportion to its importance. This makes many simple subjects appear more difficult than they really are. One example is the discussion of the arithmetic mean—an important subject, but one nevertheless easily grasped in comparison with many. Nowhere can a complaint be directed to the effect that a topic is not clearly treated, unless it is the rather confusing discussion of "discrete and continuous" data in the first chapter.

Following the discussion of the arithmetic mean is a chapter on other measures of central tendency. It is hard to appreciate the authors' distinction between "mid-measure" and "median." The mid-measure they describe as the mid-score, or in an even-numbered series, the average of the two scores at the middle. The median is the point above (or below) which half the measures lie. The description of "normal zone" is a commendable inclusion. One application would be in connection with normal weights. It is becoming recognized that there is no single weight correct for a certain age, but rather that there is a "normal zone." While some

features must be decided arbitrarily, it is nevertheless a useful concept of central tendency.

The complete treatment given percentiles and percentile ranks makes this rather easily understood subject appear difficult. Chapter 5 on measures of variability, in addition to quartile and average deviation, is an excellent presentation of that more useful measure, standard deviation.

A description of the normal probability curve follows to aid in understanding the relation of a sample to its parent distribution. The subject of reliability is given an excellent discussion with lucid explanations and examples. The chapter closes with "permutations and combinations" which has no great value to the ordinary student of statistics.

In chapter 7 the authors apply the probability curve to the distribution of class marks. They advocate T-scores as a uniform method of interpreting scores. The T-score has two advantages over percentile scores. First, a T-score of 50 always represents the mean. Second, T-scores represent on the scale of T-scores the true relative position of the raw scores on the scale of raw scores. Nevertheless, the concept of "standard deviation" (necessary to an understanding of T-scores) is less commonly appreciated than is the term "percentage."

An excellent presentation and development of correlation techniques follows in chapter 8. The next chapter, on "the interpretation and uses of correlation" is supplementary to chapter 8. In that sense it might be omitted if necessary, but the treatments of coefficient of alienation and regression equations are almost necessary, for they are often important in the practical administration of correlation techniques. The validity coefficient of a combination of tests is perhaps less widely used.

"Other measures of correlation" are well treated in chapter 10 for the benefit of the person who delves a little deeper into statistical methods. The measures here referred to are the correlation ratio, coefficient of contingency, partial and multiple correlation. Their discussion must naturally be somewhat complex for a beginning class.

GEORGE R. GRIFFITHS,
Ohio University

HUSBAND, R. W. *General Psychology.* New York: Farrar and Rinehart, Inc., 1940. Pp. xiv + 513. $2.75.

In the preface the author mentions three objectives which he has sought constantly to keep in mind: to be practical, to use the genetic approach, and to consider his subject as a social rather than a biological science. After an initial chapter in which psychology is defined in a behavioristic fashion, he has a short interesting section on genetic background in which he shows similarities and differences between men and animals, especially the apes. Then, after a brief treatment of neural and sensory processes, there follow such customary topics as motivation,

emotions, personality, learning, memory and thought; there is considerable emphasis on personality adjustments and individual differences.

Most college students should like this book; it is written from their point of view, in their terminology, and deals with a tremendous range of situations, with most of which they are already at least partially familiar from everyday life. Each topic is systematically organized, with headings and subheadings. The material is interestingly, simply and clearly presented, so that it should be easy to understand.

Some psychologists may feel that each of the very numerous topics dealt with, though well organized within itself, lacks integration with other topics, and that more general and unifying principles are subordinated to rules of thumb. Some may feel that the treatment of the subject matter is not as precise and rigorous as they would like it to be; and that terms are often used in a vague popular sense, without adequate definition.

An example of the clear organization of an interesting subtopic and of the emphasis on disparate facts and illustrations, rather than on general principles, is shown in the treatment of humor. In the four pages (193–197) devoted to this subject, seven approaches to the humorous are differentiated, described, numbered and clearly illustrated; but only one sentence is devoted to the integration of those seven approaches and little attempt is made to relate humor to other aspects of life.

As an example of the interesting, practical nature of the subject matter and of the lack of precise definition, we might turn to the treatment of the inferiority complex (266–269). Numerous types of inferiority and effective methods of dealing with them are systematically described without any attempt, however, rigorously to define an inferiority complex or to differentiate it from inferiority feelings or conflicts.

Occasionally there are statements such as the following whose accuracy or implications might be challenged. Though taken out of context, these statements can be taken as representative of the author's point of view.

The soul and consciousness are tossed aside as more or less synonymous: "There are vast areas of unexplored territory in the soul or consciousness" (5).

It is possible, of course, that the University of Wisconsin has something that other colleges don't have, but we suspect that the following statement may be questioned: "By means of a formula employing high school grades and intelligence, it is predicted with surprising accuracy [at the University of Wisconsin] just how well each student will do in college" (8).

In contrasting the "observational or empirical" methods in psychology with "common sense or rational" methods the author states: "in using this approach [common sense or rational], one considers how people

ought to behave'' (12–13). This may or may not be largely true of the rational approach, but it certainly is not entirely true of common sense.

''In the new born child almost any stimulus can elicit virtually any response'' (45). Though the behavior of the new born child is diffuse, such a sweeping statement scarcely seems warranted.

''Imagination of a high degree is another characteristic of the thought processes of children'' (63). This is probably only true of children who are encouraged to be imaginative by teachers, parents or other adults.

''Pride makes each one of us think his way of doing things is the best'' (143). This is the student's terminology, but scarcely that of a psychologist seeking to avoid an outworn faculty psychology.

''Anger and fear are both innate emotions'' (186) and are the result of stimuli which ''produce psychic shock'' (187). General emotional disturbance is commonly considered to be innate under certain circumstances, but specific emotions like fear or anger may well involve some learning.

''An instinct denotes an innate drive which persists as a general force'' (152). Most psychologists would probably want to distinguish between instincts and organic urges or drives.

For some psychologists, this book will not be acceptable as a text, even for beginners, because of a lack of unifying principles and of a sufficiently rigorous and precise treatment; for others, it will be just the interesting, practical, easily understood and comprehensive text they have been looking for.

CLARENCE LEUBA,
Antioch College

NEW BOOKS AND PAMPHLETS

Books and pamphlets for review should be sent to James P. Porter, Editor, JOURNAL OF APPLIED PSYCHOLOGY, Ohio University, Athens, Ohio.

The Adolescent Personality. PETER BLOS. D. Appleton-Century Company, New York, 1941. $3.00. 517 pp.

American College Personnel Association—Seventeenth Annual Report. KARL P. ZERPOSS, Editor. Conference at St. Louis, Mo., February 20–24, 1940. 175 pp.

The Background for College Teaching. LUELLA COLE. Farrar & Rinehart, New York, 1940. $3.50. 616 pp.

Carnegie Foundation for the Advancement of Teaching. Thirty-fifth Annual Report, 1939–40. 522 Fifth Avenue, New York City. 170 pp.

The Civilian Conservation Corps. American Youth Commission of the American Council on Education, 744 Jackson Place, Washington, D. C. 23 pp.

Counseling the Handicapped. HOLLAND HUDSON AND ROSETTA VAN GELDER. National Tuberculosis Association, 1790 Broadway, New York, 1940. 55 pp.

Directory of Applied Psychology. C. M. LOUTTIT. American Association for Applied Psychology, Inc., Indiana University, Bloomington, Indiana, 1941. 95 pp.

Farmers in a Changing World. 1940 Yearbook of Agriculture. United States Department of Agriculture, Washington, D. C., 1940. $1.50. 1215 pp.

The First Step in Guidance: Self-Appraisal. DAVID G. RYANS. Cooperative Test Service, American Council on Education, 15 Amsterdam Avenue, New York, 1941. 35 pp.

Fundamental Economic Issues in National Defense. HAROLD G. MOULTON. The Brookings Institution, Washington, D. C., 1941. 32 pp.

General Psychology. R. B. CATTELL. Sci-Art, Cambridge, Mass., 1941. $3.50. 624 pp.

Growing up in the Black Belt. CHARLES S. JOHNSON. American Council on Education, Washington, D. C., 1941. $2.25. 360 pp.

A Guide to Guidance. CHARLES M. SMITH AND MARY M. ROOS. Prentice-Hall, New York, 1941. $3.00. 44 pp.

Guideposts for Rural Youth. E. L. KIRKPATRICK. American Council on Education, Washington, D. C., 1940. 167 pp.

Hoaxes. CURTIS D. MACDOUGALL. Macmillan Company, New York, 1940. $3.50. 336 pp.

How To Be a Good Foreman. CHARLES REITELL. Ronald Press Company, New York, 1937. $1.50. 186 pp.

The Human Element in Personnel Management. LAWRENCE A. APPLEY. Society for Personnel Administration, P. O. Box 266, Washington, D. C., 1941. 35 cents. 34 pp.

Human Nature. KURT GOLDSTEIN. Harvard University Press, Cambridge, Mass., 1940. $2.50. 258 pp.

Human Nature and The Social Order. E. L. THORNDIKE. Macmillan Company, New York, 1940. $4.00. 1019 pp.

The Indivisibility of Mathematical Talent. G. REVESZ. From Acta Psychologica, Vol. V., No. 2–3. N.V. Noord-Hollandsche Uitgevers Mij, Amsterdam, 1940. 21 pp.

Latin America Educational Journals, a partial list, 1940. Division of Intellectual Cooperation, Pan American Union, Washington, D. C. 25 pp.

Let's Understand Each Other. ELEANOR R. WEMBRIDGE. The Woman's Press, New York, 1940. 322 pp.

The Measurement of Abilities. PHILIP E. VERNON. University of London Press, London, England, 1940. 10s. 6d. 308 pp.

The Measurement of Social Intelligence—A Bibliography. REIGN H. BITTNER. Ohio College Association, Bulletin No. 117, Ohio State University, Columbus, Ohio. 20 pp.

Mental Measurements Yearbook—1940. OSCAR KRISEN BUROS, Editor. Braunworth & Company, Bridgeport, Conn., 1941. $6.00. 674 pp.

National Institute of Industrial Psychology—Annual Report Ending September 1940. Aldwych House, London, England. 19 pp.

Next Steps in National Policy for Youth. American Youth Commission of the American Council on Education, Washington, D. C., 1941. 18 pp.

Occupational Counseling Techniques. WILLIAM H. STEAD, CARROLL L. SHARTLE, JAY L. OTIS AND ASSOCIATES. American Book Company, 1940. 273 pp.

Personality and Life. LOUIS P. THORPE. Longmans, Green & Company, New York, 1941. $2.50. 266 pp.

The Plans of Men. LEONARD W. DOOB. Yale University Press, New Haven, Conn., 1940. $3.00. 411 pp.

The Practice of Clinical Psychology. STANLEY D. PORTEUS. American Book Company, New York, 1941. 579 pp.

Pragmatism and Pioneering in Benoy Sarkar's Sociology and Economics. NAGENDRA NATH CHANDHURG. Chuckervertty Chatterjee & Company, Calcutta, India, 1940. 152 pp.

Principles of Abnormal Psychology. A. H. MASLOW AND BELA MITTELMANN. Harper & Brothers, New York, 1940. $3.50. 638 pp.

Progress of Scientific Research in the Field of the Exceptional Child. Proceedings of the Sixth Institute on the Exceptional Child, Child Research Clinic, Woods Schools, Langhorne, Pa., October 1939. 64 pp.

Psychodynamics of Abnormal Behavior. J. F. BROWN. McGraw-Hill Book Company, New York, 1940. $3.50. 484.

Reading and Listening Comprehension at Various Controlled Rates. HARRY GOLDSTEIN. Bureau of Publications, Teachers College, Columbia University, New York, 1940. $2.00. 69 pp.

Reading and the Educative Process. PAUL WITTY AND DAVID KOPEL. Ginn & Company, 1939. 374 pp.

Skilled Workers for Defense Industries. CHARLES S. SLOCOMBE. Personnel Research Federation, Lincoln Building, New York City, 1940. $2.00. 79 pp.

Spelling Vocabulary Study. EMMETT ALBERT BETTS. American Book Company, 1940. 72 pp.

Therapy of the Neuroses and Psychoses. SAMUEL HENRY KRAINES. Lea & Febiger, Philadelphia, 1941. $5.50. 412 pp.

The Three R's and the Adult Worker. K. E. Norris. McGill University, Montreal, 1940. $2.00. 213 pp.

Training Industrial Workers. National Institute of Industrial Psychology, Aldwych House, London, W.C. 2, September, 1940. 3 pence. 12 pp.

Units of Work and Centers of Interest in the Organization of the Elementary School Curriculum. SADIE GOGGANS. Bureau of Publications, Teachers College, Columbia University, New York, 1940. $1.60. 140 pp.

Wisdom for Widows. TONI TORRY. E. P. Dutton & Company, New York, 1941. $2.50. 252 pp.